KYAMBOGO UNIVERSITY

MATHEMATICAL COMPUTATIONS, STATISTICS AND PROBABILITY

This course leads to the award of Diploma in Education Primary of Kyambogo University.

Kyambogo University

Authors:

David Nyakairu	:	Kyambogo University
Samuel Sebabi	:	Kyambogo University
Sarah Kisa	:	Kyambogo University
Jane Kibuuka	:	Kibuli PTC

Reviewer (s)

David Nyakairu	:	Kyambogo University
Samuel Sebabi	:	Kyambogo University
Sarah Kisa	:	Kyambogo University
Jane Kibuuka	:	Kibuli PTC

Editors:

David Nyakairu	:	Kyambogo University
Sarah Kisa	:	Kyambogo University

Design and Production Team:

Mr. Aron Y. Otto	:	Ag. Head of Distance Education Department
Mr. Andrew Meya	:	Assistant Coordinator, DEPE
Mary Lilian Ziraba	::	Typesetter

ACKNOWLEDGEMENTS

Kyambogo University is grateful to the authors, reviewers and publishers for the reference materials consulted in the production of this self-study module.

Above all, the Distance Education Department is also grateful to the writers, editors, and graphic designer and design team for the good work done to produce this module within the limited time available.

We appreciate your support to this programme and encourage you to continue doing so.

TABLE OF CONTENTS

GENERAL INTRODUCTION

Dear Student,

You are most welcome to this second module in Mathematics (MT/2). This is the second module in the series of Mathematics modules you will receive during your course.

In this module, you will be introduced to:

- Mathematical computations.
- Statistics
- Probability

Unit 1: It deals with mathematical computations. You will study various ways of how to carry out computations in mathematics.

Unit 2: It covers statistics. You will learn various ways of presenting data. You will also learn how to find measures of central tendency, dispersion and their applications.

Unit 3: It covers probability. You will learn probability theory using set theory and tree diagrams.

Note that:

Throughout these units you will find self-check activities to do. remember that these activities are meant to make you study more effectively. Please attempt all the activities as they will support your learning. You will find that the activities are much useful if you do them as you read the unit. They are not only reinforcing the instruction but are also planned to motivate you to learn. Do not skip them or leave them until the end of the unit.

When you finish an activity, turn to the end of the unit and check the possible answers. If you find that your answer is not correct, study the unit again and try to see where you went wrong. Discuss any problem you meet with other students and colleagues – they may be able to help you. If the reasoning is still not clear, make an accurate note of the difficult part and seek advice from any resource persons, you find in your area, or better still, present the problem to your facilitator during face-to-face residential session.

So you now know this module is organized to help you learn. The other self-study modules of the course are organized the same way.

We wish you all the best as you work through this module.

SYMBOLS USED

Throughout each unit, a number of symbols are used. Some indicate that you should do something. Others indicate things like the learning outcomes of the unit. These symbols are guides that will show you where you are and what to do.

The symbols and the meanings are these:

Outcomes: This symbol indicate the learning objectives and expected outcomes of the unit, including benefits to you. If you do what is learnt.

Activity: The symbol indicates that you should complete the activity indicated before you proceed with the reading.

Note this important point: This tells you to take note or to remember an important point.

A question: This indicate that there is a question that you should answer or think about.

Further reading: This indicates additional reading from another source or module that is suggested.

✓

Checking: This symbol is used to indicate that you are required to check through something, either a piece of work you have just written or a list of points to be considered in the text.

Summary: This indicates that a summary is given of what has been covered in the unit.

Discuss: This indicates activities in which you are asked to discuss ideas with others.

Hand in: This shows something that should be handed in to a facilitator.

Congratulations: This shows that you have really achieved something! When you see this you will know that you have achieved an important point in the learning programme.

UNIT 1

MATHEMATICAL COMPUTATIONS

INTRODUCTION

Dear Student,

You are welcome to unit 1 of module 2 that is entitled Mathematical Computations.

AIMS

The aims of the unit are:

1. To introduce you to how to simplify surds and indices.
2. To enrich your knowledge of sequence and introduce you to series.
3. To strengthen your knowledge of algebraic expressions.

SPECIFIC OBJECTIVES

By the time you have gone through this unit, you should be able to:

1. Simplify expressions that contain surds and indices.
2. Write down sequences.
3. Identify the first term, common differences
4. Find the sum of a given number of terms in a series.
5. Expand and factorize algebraic expressions.

UNIT ORGANISATION

This unit is organized into 4 topics.

Topic 1: Surds, indices and logarithms has three sub-topics:

(a) Surds
(b) Indices
(c) Logarithms

Topic 2: Sequence and series; have no sub-topics

SUBJECT ORIENTATION

Before you begin studying this unit, you should revise number work and algebra that were covered in Module MT/1. You should also be familiar with writing numbers in standard/scientific form.

❒ STUDY REQUIREMENTS

You will need to have a four-figure table, pen, pencil, ruler and an exercise book. You also need to be in a quite room free from internal and external disturbance.

Enjoy learning more about Mathematical computations.

TOPIC 1: SURDS, INDICES AND LOGARITHMS

(a) **SURDS**

(i) What is a surd?

A surd is an irrational number, Irrational numbers have no exact value. Any ROOT, which has no exact value, is known as a surd.

Now consider the following cases:

Case 1:

$$\sqrt{4} = 2, \qquad \sqrt[3]{27} = 3, \qquad \sqrt[4]{16} = 2,$$

$$\sqrt{6\frac{1}{4}} = \sqrt{\frac{25}{4}} = \frac{5}{2}$$

$$\sqrt{0.009} = .03 = \frac{3}{10}$$

Case 2:

$$\sqrt{2} = 1.414213 \ldots\ldots\ldots; \qquad \sqrt{5} = 2.236067\ldots\ldots\ldots$$

$$\sqrt[4]{64} = 2.828427$$

We observe that:

In case 1, the numbers under the root sign are positive rational numbers and the roots of these numbers are either integers or can be expressed as a ratio of two integers.

A number which is either an integer or can be expressed as a fraction of two integers when the denominator is not zero is called a rational number.

If a number cannot be expressed as an integer, a non-terminating decimal or a recurring decimal it is called an irrational number.

Now consider case 2. The numbers under the root sign are positive rational numbers but the roots of these numbers are irrational. Such numbers are called surds.

For example:

$\sqrt{16}$ = 4 is not a surd but $\sqrt[3]{16}$ is a surd,

$\sqrt{8}$ is a surd but $\sqrt[3]{8}$ = 2 is not a surd.

Determine whether or not the following roots are surds.

(i) $\sqrt{36}$ (ii) $\sqrt[3]{64}$ (iii) $\sqrt{10}$ (iv) $\sqrt[3]{8}$ (v) $\sqrt[3]{32}$

(ii) **Order of Surds**

$\sqrt{2}$, $\sqrt{5}$, $\sqrt{11}$ and so on, are surds of order 2.

$\sqrt[3]{2}$, $\sqrt[3]{5}$, $\sqrt[3]{30}$ and so on, are surds of order 3.

$\sqrt[4]{2}$, $\sqrt[4]{20}$, $\sqrt[4]{64}$ and so on, are surds of order 4.

In this topic will consider mainly surds of order 2.

(iii) Manipulation of Suds

1. Addition and Subtraction

Terms which contain **<u>LIKE</u>** surds can be added or subtracted, for example:

$2\sqrt{3} + 3\sqrt{3} = 5\sqrt{3}$

$\sqrt{5} + \sqrt{20} = \sqrt{5} + \sqrt{4x5} = \sqrt{5} + \sqrt{4x}\ \sqrt{5} = \sqrt{5} + 2\sqrt{5} = 3\sqrt{5}$

$\sqrt{5}\,4 + \sqrt{150} - \sqrt{24} = \sqrt{9x6} + \sqrt{25x6} - \sqrt{4x6} = 3\sqrt{6} + 5\sqrt{6} - 2\sqrt{6} = 6\sqrt{6}$

Terms which contain **<u>UNLIKE</u>** surds cannot be added or subtracted, for example:

$2\sqrt{3} + 7\sqrt{2}$ cannot be simplified any further.

[*NB* : Compare this rule with the rule for algebra terms such as 2a; 3a, 7b, 5b…….]

Add or subtract the following where possible.

(i)	$\sqrt[4]{3} - \sqrt{3}$	(vi)	$\sqrt{20} + \sqrt{5}$
(ii)	$5\sqrt{7} + 7\sqrt{7}$	(vii)	$4\sqrt{3} - \sqrt{12}$
(iii)	$12\sqrt{2} - 6\sqrt{3}$	(viii)	$5\sqrt{7} - \sqrt{28}$
(iv)	$20\sqrt{2} - 6\sqrt{2}$	(ix)	$\sqrt{11} + \sqrt{44} - \sqrt{99}$
(v)	$10\sqrt{3} + 5\sqrt{7}$	(x)	$\sqrt{18} - \sqrt{32} + \sqrt{50}$

You should have found that:

(i), (ii), (v), (vi), (viii), (ix) are possible.

Answers should be: $3\sqrt{3}$, $12\sqrt{7}$, $14\sqrt{2}$, $3\sqrt{5}$, $2\sqrt{3}$ $3\sqrt{7}$

0, $4\sqrt{2}$

Multiplication and Division

Numbers which are under the **<u>SAME</u>** root may be multiplied or divided as follows:

$$\sqrt{2x}\ \sqrt{3} = \sqrt{2x3} = \sqrt{6}$$

$$\frac{\sqrt{6}}{\sqrt{2}} = \sqrt{\frac{6}{2}} = \sqrt{3}$$

Brackets are multiplied using the Distribution Law (as for algebraic term) for example:

$$(2 + \sqrt{3})(3 - 2\sqrt{3}) = 6 + 3\sqrt{3} - 4\sqrt{3} - (2x3) = -4\sqrt{3}$$

Rationalization of a denominator is the usual practice i.e. no surd should be left in the denominator of a function.

- A function which has a simple surd in the denominator has both numerator and denominator multiplied by that surd e.g. $\frac{1}{\sqrt{3}}$

To rationalize the denominator multiply numerator and denominator by $\sqrt{3}$, to get.

$$\frac{1x\sqrt{3}}{\sqrt{3x\sqrt{3}}} = \frac{\sqrt{3}}{3} =$$

- A function which has a two-term denominator has its denominator and numerator multiplied by the complementary bracket as follows e.g.

- To rationalize $\frac{1}{2+3\sqrt{3}}$ multiply numerator and denominator by

  (2 - 3 $\sqrt{3}$) to obtain:

$$\frac{1}{(2+3\sqrt{3})} \text{ x } \frac{(2-3\sqrt{3})}{(2-3\sqrt{3})} = \frac{(2-3\sqrt{3})}{[4-(9x3)]} = -\frac{(2-3\sqrt{3})}{23}$$

- This type of fraction uses the factor property known as "difference of two squares" i.e. $a^2 - b^2 = (a - b)(a + b)$,

  Now see whether you can rationalize the denomination of the following:

(i) $\frac{2}{\sqrt{2}}$ (ii) $\frac{3}{\sqrt{7}}$ (iii) $\frac{2\sqrt{3}}{\sqrt{6}}$

(iv) $\frac{1}{(2-\sqrt{3})}$ (v) $\frac{(2+\sqrt{3})}{(3-3\sqrt{3})}$ (vi) $\sqrt{27}$ x $\sqrt{15}$

(vii) $\frac{20}{5\sqrt{20-2\sqrt{45}}}$

You should have the following answers:

(i) $2\frac{\sqrt{5}}{5}$ (ii) $\frac{3\sqrt{7}}{7}$ (iii) $\sqrt{2}$

(iv) $14 + 7\sqrt{3}$ (v) $\frac{-1}{3}(12 + 7\sqrt{3})$ (vi) $9\sqrt{6}$

(vii) $\sqrt{5}$

3. **Simplification**

Simplification of surds combines two or more of the above processes.

Examples:

1. Remove from the root any factor whose root is an exact number e.g.

- $\sqrt{12} = \sqrt{4x3} = 2\sqrt{3}$
- $\sqrt{72} = \sqrt{32x2} = 6\sqrt{2}$

You should then have the smallest possible number under the root sign.

Combine any terms which have LIKE surds and can be added or subtracted e.g. if you are asked to simplify:

$\sqrt{12} + \sqrt{36} + \sqrt{27},$ proceed as follows:

$\sqrt{12} = 2\sqrt{3}$

$\sqrt{36} = 6$

$\sqrt{27} = \sqrt{9x3} = 3\sqrt{3}$

Hence $\sqrt{12} + \sqrt{36} + \sqrt{27} = 2\sqrt{3} + 6 + 3\sqrt{3} = 5\sqrt{3}$

If you are asked to simplify a fraction, cancel anything which will cancel and rationalize the denominator if necessary e.g.

$$\frac{6+5\sqrt{3})}{2+10\sqrt{3}} = \frac{1}{2} \text{ since } \qquad \frac{6+5\sqrt{3}}{12+10\sqrt{3}} = \frac{(6+5\sqrt{3})}{2(6+5\sqrt{3})} = \frac{1}{2}$$

ACTIVITY: MT/2/1-1

1. Express the following in terms of $\sqrt{2}$

   (i) $\sqrt{18}$ (ii) $\sqrt{20}$ (iii) $\sqrt{200}$ (iv) $\sqrt{7200}$

2. Rationalize the denominators of the following:

   (i) $\frac{2}{\sqrt{3}}$ (ii) $\frac{3}{\sqrt{5}}$ (iii) $\frac{5}{\sqrt{7}}$ (iv) $\frac{2}{\sqrt{11}}$

3. (i) $\frac{1}{(\sqrt{2}-1)}$ (ii) $\frac{1}{(\sqrt{3}-\sqrt{2})}$ (iii) $\frac{(\sqrt{5}+\sqrt{3})}{(\sqrt{5}-\sqrt{3})}$

4. Express $\sqrt{5000}$ in terms of the simplest and possible.

5. Remove the brackets from: $(2\sqrt{5} + 1)(3\sqrt{5} - 1)$

6. Without using tables, find the value of: $(3\sqrt{5} - \sqrt{2})(3\sqrt{5} - \sqrt{2})$

(b) **INDICES**

(i) Indices

1. A number written as a power of another is said to be written in INDEX form e.g. 3^4 which is a shortened way of expressing 3 x 3 x 3 x 3. In this example the INDEX (a POWER) in this case is 4, the BASE number is 3. The plural of the word index is INDICES.

   (i) Write in index form: 3 x 3 x 3; 5 x 5 x 5 x 5
   (ii) Evaluate 2^5 ; 7^3

   [You should have 3^3, 5^4 as answers for (i) and 32, 343 as answers for (ii)]

2. If two or more numbers are written in index form, using the SAME base, they can be multiplied or divided by using the laws of indices.

N.B: Numbers written in index form CANNOT be ADDED or SUBTRACTED in that form.

(i) Suppose we want to calculate $3^4 \times 3^3$. One way of working would be to write 3^4 as $3 \times 3 \times 3 \times 3$ and 3^3 as $3 \times 3 \times 3$.

The product of 3^4 and 3^3 can be written as $3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3$ which is equivalent to 3^7. What do you notice about the indices of the numbers involved?

(ii) Similarly if we want to calculate $3^4 \div 3^3$, we can also use the expanded form as follows: $\dfrac{3 \times 3 \times 3 \times 3}{3 \times 3 \times 3} = 3^1$

What do you notice about the indices of the numbers involved in this case?

These examples illustrate the multiplication and division laws of indices i.e.

1. When multiplying two numbers in index form, the product is found by <u>ADDING</u> the indices.

2. When dividing two numbers in index form, the quotient is found by <u>SUBTRACTING</u> the index of the denominator from the index of the numerator.

Calculate the following, leaving your answers in index form:

(i) $2^3 \times 2^5$
(ii) $7^6 \div 7^2$
(iii) $10^5 \times 10^2$
(iv) $12^{10} \div 12^5$

The base does not have to be a number, but may be a letter e.g.

$a^8 \times a^5 = a^{13}$, $b^{10} \div b^2 = b^8$, $X^2 \times X^3 = X^5$ and so on.

[You should have 2^8, 7^4, 10^7, 12^5 as answers for (i) - (iv)

(ii) **The meaning of the Zero Index:**

Using the laws above, calculate the following:

(i) $2^3 \times 2^0$ (ii) $3^5 \times 3^0$ (iii) $5^2 \div 5^0$ (iv) $a^3 \times a^0$

Did you find that 2^3 remained as it was, 3^5 remained as it was and 5^2, a^3 as they were?

The only number which has the property of multiplying or dividing and giving the original term as the answers is 1. Therefore $2^0 = 1$, $3^0 = 1$, $5^0 = 1$.

In this case, we have established that anything to the power zero is 1.

(iii) **The meaning of Negative Indices**

Using the laws of indices, calculate the following:

(i) $2^5 \times 2^{-1}$ (ii) $7^2 \times 7^{-2}$ (iii) $2^5 \times \frac{1}{2}$

(iv) $7^2 \times \frac{1}{7^2}$

Did you find that (i) and (iii) both came to 2^4 and (ii) and (iv) both came to 7^0 or 1?

Now find the value (in algebraic terms) of:

(i) $a^7 \times a^{-2}$ (ii) $a^7 \div a^{+2}$ (iii) $a^7 \times \frac{1}{a^2}$

Did you find that they all gave the result a^5?

These examples illustrate the meaning of a negative index e.g.

$2^{-1} = \frac{1}{2}$ $7^{-2} = \frac{1}{7^2}$ $a^{-2} = \frac{1}{a^2}$

In general terms, a number with a negative index is equivalent to the reciprocal of the same base number with the positive power of the same denominator.

(iv) The meaning of Fractional Indices

Using the laws of indices, calculate the following:

(i) $2^{½} \times 2^{½}$ (ii) $a^{½} \times a^{½}$ (iii) $3^{½} \times 3^{½} \times 3^{½}$

Did you find the answers as 2^1, a^1, 3^1?

These examples illustrates the fact that fractional indices represents roots e.g. $2^{½}$ means

$\sqrt{2}$, $a^{½} = \sqrt{a}$, $3^{½}$ means the cube root of 3 and so on. Fractional indices can have a numerator greater than one e.g. $a^{½}$ which means (the cube root of a square or $(\sqrt[3]{a})^2$

The denominator indicates a root, the numerator indicates a power. Evaluate the following:

$\left(\frac{9}{64}\right)^{\frac{1}{2}}$, $\qquad 8^{\frac{2}{3}}$, $\qquad 100^{\frac{3}{2}}$

[You should have $\frac{3}{8}$, $2^2(= 4)$, 10^3 (= 1000) as answers]

Fractional indices can also be negative, in which case the two meanings have to be combined e.g. $a^{\frac{-1}{3}}$ means $\frac{1}{3\sqrt{a}}$; $9^{\frac{-3}{2}}$ means $\frac{1}{(\sqrt{9})^3} = \frac{1}{3^3}$

Evaluate the following $\left(\frac{9}{64}\right)^{\frac{1}{2}}$ $\quad 8^{\frac{2}{3}}$ $\qquad 100^{\frac{3}{2}}$

[You should get $\frac{8}{3}$, $\frac{1}{4}$, $\frac{1}{1000}$ as answers]

N.B: There are two ways of trying to evaluate something like $100^{\frac{3}{2}}$ (for example)

One may write it as $(\sqrt{100})^3$, the other may write it as $\sqrt{(100)^3}$
The former is usually easier because it involves smaller numbers. If you are using a calculator both are equally straightforward.

ACTIVITY: MT/2/1-2

Do not use a calculator.

1. Calculate the following, leaving your answers in index form.

   (i) $2^3 \times 2^5$
   (ii) $3^7 \times 3^3$
   (iii) $3^2 \times 3^{-2}$
   (iv) $6^3 \times 6^0$
   (v) $a^5 \times a^{-1}$

2. Evaluate the following:

   (i) 4^{-2}

   (ii) $216^{-⅔}$

   (iii) $\left(\frac{2}{3}\right)^0$

   (iv) $\left(8^{-⅓}\right)^3$

   (v) $\left(\frac{169}{144}\right)^{-½}$

3. Evaluate: $\dfrac{27^{½} \times 144^{3/2} \times 13^2}{169^{½} \times 156}$

✓ Check your answers with those given at the end of this unit.

REVISION

Before going on to logarithms, it is necessary to revise standard form significant figures and decimal places.

(a) Standard form (sometimes called Scientific Notation)

To write a number in standard form, it must be written in the form $a \times 10^n$ where a is a number between 1 and 10, and n is an integer, for example:

103 in standard form is 1.03×10^2
5678 in standard form is 5.678×10^3
9.63 in standard form is 9.63×10^0
0.845 in standard form is 8.45×10^{-1}

0.03129 in standard form is 3.129×10^{-2} and so on.

Write the following numbers in standard form.

15.2, 235, 0.112, 0.0378, 0.1172, 0.001172

For the first, $1.52 = 1.52 \times 10^{1}$

[Similarly the others are: 2.35×10^{2}, 1.12×10^{-1}, 3.78×10^{-2}, 1.172×10^{-1}
1.172×10^{-3}]

Standard form is extremely helpful when dealing with computation using logarithm tables.

(b) Significant Figures

Significant figures are concerned with the degree of accuracy required for a number usually the answer to a calculation. For instance, if we use four – figure tables to do a piece of computation, we will be expected to give the answer correct to 3 significant figures. If the answer from the tables was 25.43, then correct to 3 significant figures, it would be 25.4. if the answer was 25.47, correct to 3 significant figures, it would be 25.5. The question we ask ourselves is "Is the answer nearer to 25.4 or 25.5? We can judge this by looking at the fourth figure. In the first case it is 3, which is less than 5, so the number is nearer to 25.4; in the second case the fourth figures is 7, which is greater than 5, so the number is nearer to 25.5. If the fourth figures was 5, there would be a choice between 25.4 and 25.5, but by convention we usually take the higher figure i.e. 25.5.

Correct the following numbers to 3 significant figures: 73.21, 101.6, 1237, 0.7891, 0.03587, 15, 390.

[First ask yourself the question, "Is it near to? or to?] [See the end of the section for answers].

* In the case of 0.7891 and 0.03587, the zeros are NOT significant figures, since they are simply indicating the place of the point (giving the degree of the number). Similarly the zero at the end of 15390 indicates the degree (size) of the number i.e. fifteen thousand rather than one thousand, five hundred….*

(c) Decimal Places

Decimal places are also concerned with the degree of accuracy of a number. The same process is used as significant figures, but in this case it is only the figures after the decimal point which are counted.

<u>For example</u>: 1,723 corrected to 2 decimal places is 1.72

101.456 corrected to 2 decimal places is 101.46
0.03457 corrected to 3 decimal places is 0.035

☞ **Note**: that in this case, zeros after the decimal point are counted.

Correct the following numbers to 2 decimal places: 105,432, 2,678, 0.0692

Answer: To 2 decimal places: 105,43, 2.68, 0.07

(d) Logarithms

(i) Logarithms are indices

When a number is written in index form e.g. 3^4, then the logarithm of 3^4 (or 81) to the base 3 is 4. Similarly the logarithm of 100, (witch is 10^2), to the base 10, is 2.

In general: if $a^x = b$
Then log b = x

Write down the logarithms of the following to the given bases.

4^2 to the base 4
3^{10} to the base 3
1000 to the base 10
a^3 to the base a

[You should have the answers: 2, 10, 3, 3]

In elementary mathematics, logarithms can be used to aid computation. By using the rules of indices, multiplication and division can be done by adding and subtracting.

For example:

$10^2 \times 10^5 = 10^7$

or logarithm of 100 + logarithm of 100,000 = logarithm of 10,000,000

Hence 100 x 100,000 = 10,000,000

This example is trivial, because it can easily be done mentally. However, Mathematicians used this principle to help in computation before the days of calculators!

(ii) **Common Logarithms**

These are all worked out to base 10, and are usually found in either three-figure of four figure forms. In order to find the logarithm of a number, first work out the

<u>CHARACTERISTIC</u> (the number before the decimal point in a logarithm), e.g. a number between 100 (10^2) and 1000 (10^3) will have a logarithm 2. Its characteristic is 2. A number between 10 (10^1) and 100 (10^2) will have a logarithm 1. Its characteristic is 1.

If a number is written in standard form, its characteristic may be determined from the power of 10, e.g. 11.35 in standard form is 1.135×10^1, and the logarithm of 11.35, which is between 10^1 and 10^2, has a characteristic 1.

(iii) Use of four-figure logarithm tables

Suppose we want to look up the logarithm of a four-figure number e.g. 23.54

1. First write the number in standard form i.e. 2.354×10^1

2. Identify the characteristic of the logarithm (from the power of 10) – in this case it will be 1.

3. The first two digits of the number (i.e. 23) are used to look up in the left-hand column of the table of logarithms. This then gives us the row to consider.

4. Look along the row of 23 to the column headed 5 (the third digit of the number). In this case it gives 3711.

5. Look along the same row to the columns on the far right (called the "difference" columns) under the column headed 4 (the fourth digit of the number). This gives 7.

6. Add the 7 to 3711, giving 3718.

7. The complete logarithm (to 4 figures) of 23,54 may then be written as 1.3718.

N.B: The part of the logarithm after the point is called the MANTISSA

Now look up the logarithm for each of the numbers below by using the above procedure:

1,371, 278.3, 7912, 50.63

Answers: 0.1370, 2,4445, 3,893, 1,7045

(iv) Use of Logarithms in Computation

Suppose we want to work out 1,371 x 278.3. If we have no calculator, but we have access to four-figure tables, we can look up the logarithms of the numbers. Using the rules of indices we can obtain the logarithm of the product as follows:

NO	Standard Form		Log	
1.371	1.371×10^0	$\longrightarrow$	0.1370	When multiplying ADD the
278.3	2.783×10^2	$\longrightarrow$	$^{+}\underline{2.4445}$	logarithms
			2.5815	

Now it is necessary to look back the logarithm 2.5815 in the <u>ANTI-LOGARITHM</u> tables. The characteristic of the logarithm of the product is 2. This gives us the power of 10 of the number. The mantissa 5815 will give us the digits. Anti-logarithm tables are used in the same way as logarithms i.e.

1. Look for 58 in the left-hand column, giving the row.
2. Look along this row under column 1; this gives 3811
3. Look in the difference columns under 5; this gives 4
4. Add the two numbers giving 3815
5. The complete anti-logarithm in standard form is 3.815×10^2, or 381.5

Hence $1.371 \times 278.3 = 381.5$

Correct to 3 significant figures the answer is 382. Using logarithms, evaluate the following correct to 3 significant figures:

1. 1.371×7912; (ans. 10800)

2. 278.3×50.63; (ans. 14.100)

3. $7912 \div 50.63$; (ans.. 156)

4. $2783 \div 1.371$; (ans. 203)

(v) Numbers less than 1

The characteristic of logarithms of numbers less than 1 are NEGATIVE.
Consider for example an exact power of 10 such as 0.1

Another way of writing 0.1 is 1×10^{-1} (standard form). The logarithm (base 10) of 0.1 is therefore -1. What are the logarithms of the following exact powers of 10?

0.01,0.001, 0.0001, ……………
[-2, -3, -4 and so on]

Take for example, a number between 0.1 and 0.001; 0.08532, how can we find its logarithm?

The procedure is as before.

1. Write the number in standard form i.e. 8.532×10^{-2}
2. Identify the characteristic – in this case -2.
3. Look up the digits in logarithm tables as before.

Row 85 under column 3 gives 9309, the difference column headed 2 gives 1, therefore the complete Mantissa in 9310.

We have to decide whether the Mantissa is also negative.
The number 0.08532 is between 0.01 and 0.1.

Its logarithm is therefore greater than -2 and less than -1. The Mantissa must therefore be POSITIVE. The complete logarithm is -2 + 0.9310.
We usually write this as 2.9310 and refer to the negative 2 as "bar 2".

N.B: When using logarithm tables for calculations involving numbers less than 1, it is important to remember that the characteristic of such a number is NEGATIVE, but the MANTISSA is POSITIVE.

<u>Example 1</u>: Solve using logarithms: 0.1539×0.08532

No	**Standard Form**		**Log**	**ADD**
0.1539	1.539 x 10-1	→	1.1872	(the 1 carried over
0.08532	8.532×10^{-2}	→	$^{+}2.9310$	from adding the
				mantissa is positive, 1 and
	1.313×10^{-2}	←	2.1182	2 are both negative)

Ans: (correct to 3 significant figures) will be 1.31 x 10-2 or 0.0131

Example 2: $0.1539 \div 0.08532$

No	**Standard Form**		**Log**	**Subtract**
0.1539	1.539×10^{-1}	→	1.1872	
0.08532	8.532×10^{-2}	→	-2.9310	the 1 'borrowed' other
	1.804×10^{0}	←	0.2562	than 'paid back' is
				positive)

Ans. (Correct to 3 significant figures) will be 1.80×10^{0} or 180

ACTIVITY: MT/2/1-3

(Use logarithm tables NOT a calculator)

Evaluate the following using logarithm (4 – figure) tables and give your answers correct to 3 significant figures.

1. 256.3×57.91
2. $3012 \div 189.4$
3. 3.892×0.5416
4. $3.892 \div 0.5416$
5. $\dfrac{3012 \times 189.4}{256.3 \times 0.5416}$

✓ Check your answers with those given at the end of this unit.

TOPIC 2: SEQUENCE AND SERIES

(a) Simple Number Pattern

Consider the following sets of numbers.

(i) 2, 4, 6, 8, 10, ………………
(ii) 2, 4, 8, 16, 32, ………………..
(iii) ½, ⅓, ¼, 1/5, 1/6 ……………
(iv) 1, 3, 5, 7, 9, ……………………

Find the relation between any number in the given set and the number immediately before it. Take a close observation for each set. Hopefully you will find that every number in a particular place of a given set, has some relation with the position of that place.

In set (I)

1st term 2 is 2×1
2nd term 4 is 2×2
3rd term 6 is 2×3
nth term is $2 \times n$

In set (II)

1st term 2 is 2^1
2nd term 4 is 2^2
3rd term 8 is 2^3
nth term is 2^n

In set (iii)

1st term $\frac{1}{2}$ is $\frac{1}{1+1}$

2nd term $\frac{1}{3}$ is $\frac{1}{2+1}$

3rd term $\frac{1}{4}$ is $\frac{1}{3+1}$

nth term is $\frac{1}{n+1}$

In set (iv)

1st term 1 is $2 \times 1 - 1$

2nd term 3 is $2 \times 2 - 1$

3rd term 5 is $2 \times 3 - 1$

nth term is $2 \times n - 1 = 2n - 1$

Thus each of these sets is formed, in a definite order and there is a simple rule by which the terms are indicated.

(b) Sequences

A set of numbers, stated in a definite order, such that each number can be obtained from the previous number according to some rule, is a sequence. Each number of the sequence is called a term.

Consider the following:

3, 5, 7, 9, 11, ………………..
1, 4, 9, 16, 25, ………………
1, 2, 4, 8, 16, ……………….

Each of these is sequence. The "….." at the end of each sequence show that each one could go on indefinitely (the sequence is infinite). However, if we wished to have a limited number of terms of a sequence, we could write 3, 5, 7, 9, 11,……47. The final single full stop shows that the sequence ends when the number 47 is reached. Such a sequence is said to be finite.

An explanation for the nth term (written U_n) of a sequence is useful since any specific term of the sequence can be obtained from it. The nth term of the sequence 1, 4, 9, 16, 25, ……… is n^2.

Thus for $n = 1$, first term written as $U_1 = 1^2 = 1$
$n = 2$, second term is $U_2 = 4$
$n = 3$, third term is u3 $= 9$, etc.

Example Consider the nth term of the sequence as $n^2 + 1$ and write down its first 3 terms and the 10th term.

Solution To get 1st term, put $n = 1$ then
1st term $= 1^2 + 1 = 2$
2nd term $= 2^2 + 1 = 5$
3rd term $= 3^2 + 1 = 10$
10th term $= 10^2 + 1 = 101$

In this way we can write down any term of the sequence.

Write down the first 3 terms of the sequences whose general nth terms are given below:

(i) $1/n$ (ii) $\dfrac{n - 1}{n}$ (iii) $(^-1)^n$

(iv) $3 \times 2^{n-1}$ (v) $n^2 + n - 1$

Solution:

(i) 1, ½, ⅓ (ii) 0, ½, ⅓ (iii) -1, 1, -1

(iv) 3, 6, 12 (v) 1, 5, 11

(c) **Series**

If the terms of a sequence are considered as a sum, the expression is called a series.

Example: 1 + 2 = 4 + 8 + 16 + 32 + ……………is a series.

A series may end after a certain number of terms. In this case it is called a finite series. A series, which continues indefinitely, is called an infinite series.

Example: 0.1 + 0.01 + 0.001 + 0.0001 = 0.1111 is called a finite series.

0.1 + 0.01 + 0.001 + 0.0001 + ………………
(= 0.1111……….. = 0.1 = 1/9) is an infinite series.

In case the sum of the series is needed then this can be determined. If we know the number of terms required e.g. the first ten terms which would be.

1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10

(i) **Arithmetic Progressions**

Another name commonly used for a series is PROGRESSION. The series of natural numbers above is an example of an ARITHMETIC PROGRESSION (abbreviated to A.P) because the next term of the progression is found by adding (a subtracting) the SAME number each time. In the example above the next term is found by adding ONE. In this case therefore, we say that the COMMON DEFERENCE of this A.P. is +1. In this case a1: the FIRST term is 1, the SECOND term is 2,…..and so on. We use the letter "a" for the first term in the general form and the letter "d" for the common difference. Any A.P. therefore is of the form a, a + d, a + 2d,….. Suppose we want a GENERAL term (i.e. the n^{th} term). Looking at the pattern, we see that the nth term can be written in the form a + (n-1)d.

[Check the second (n = 2), third (n = 3) and so on]

Determine whether the following sequences are A, P's if they are, write down the first term, the common difference, and hence the n^{th} term.

(i) 3, 6, 9, 12, 15, …………………..
(ii) 5, 3, 1, -1, -3, ……………………
(iii) $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}$, ………………………

[You should have decided that (i) and (ii) are A.Ps]
For (i), a = +3 t therefore n^{th} term = 3 + (n – 1).3.

This can be simplified to 3 + 3n - 3

For (ii) a = 5, d = -2 therefore n^{th} term = 5 + (n-1) (-2)
This can be simplified to 5 – 2n + 2 = 7 – 2n.

When the sum of n terms of an A.P is required a formula can be developed as follows:

Let the sum of n terms:
S_n = a + 9a + d) + (a + 2d) + ……+(a + (n – 1)d)……………..(1)

Write the terms backwards:
(Sn) = (a + (n – 1)d) + a +(n – 2)d) + (a + (n – 3)d) + ….+ (a + 2d) + (a + d) + a ….. (2)

Adding (1) and (2)

$$2(S_n) = 2na + n\,[9n - 1)d]$$
$$2S_n = n\,[2a + (n - 1)d]$$
$$S_n = \frac{n}{2}\,[2a + (n - 1)d]$$

Now find the sums of the following:

(i) 10 terms of the A.P. 1, 2, 3, 4, ………………..

[In other words the first ten natural numbers]

(ii) 20 terms of the A.P. 3, 6, 9, 12, ……………….

(iii) 12 terms of the A.P. 5, 3, 1, ……………………

[You should have answers (i) 55, (ii) 630, (iii) -72]

(ii) **Geometric Progressions**

A GEOMETIRC Progression (G.P) is formed by multiplying one term by the same number in order to get the next. An example of a G.P. is 1, 5, 25, 125,….. In this case the first term is 1 and the multiplying factor or COMMON RATIO (r) is 5.

The general form of a G.P. is a, ar, ar^2 ………………..
Then n^{th} term is ar^{n-1}. [Check the 2^{nd} and 3^{rd} terms]
Determine whether the following sequence are G.Ps. If hey are, write down the first term, the common ratio and hence the nth term.

(i) 4, 12, 36, 108 ……………………

(ii) 1, -1, $\frac{-2}{2}$ …………………………..

(iii) 1, $\frac{-1}{2}$ $\frac{1}{4}$ $\frac{-1}{8}$ $\frac{1}{16}$..................

[You should have decided that (i) and (iii) are G.P's]

For (i), $a = 4$, $r = 3$. therefore nth term is $4(3)^{n-1}$

For (iii) $a = 1$, $r = \frac{-1}{2}$. therefore nth term is $1\left(\frac{-1}{2}\right)^{n-1}$

In this later case, r is both negative and less than 1, therefore the terms are decreasing in numerical value (modules) and alternating in sign.

When the SUM of n terms of a G.P is required, a formula can be developed as follows:

Let the sum of terms $S_n = a + ar + ar^2 + \ldots + ar^{n-2} + ar^{n-1}$………..(1)

Multiply throughout by r: $rS_n = ar + ar^2 + ar^3 + \ldots + arn^{-1} + ar^n$ ….(2)

Subtracting (2) from (1)

$S_n - rS_n = a - ar_n$

$Sn (1 - r) = a(1 - r^n)$

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

The first term is more convenient if r < 1, and the second if r > 1.
Now find the sums of the following, leaving your answer in index form.

(i) 10 terms of the G.P. 1, 5, 25 …………..
(ii) 20 terms of the G.P. 4, 12, 36, …………
(iii) 12 terms of the G.P. 1, $\frac{^-1}{2}$, $\frac{1}{4}$, ………..

Answer should be as follows:

(i) $\frac{1 (5^{10} - 1)}{4}$

(ii) $2(3^{20} - 1)$

(iii) $[2(1 - \frac{1}{2^{12}})]$

(d) **Simple and Compound Interest**

The difference between simple and compound interest (e.g. on a saving account) is the difference between an A.P. and a G.P.

Take an investment of Shs. 10,000 for 5 years, at 25. With simple interest the amount at the end of 5 years will be Shs. 10,000 + Shs. 5 x 200 = shs.11,000. This could be written as Shs. 10,000 + 200 + 200 + 200 + 200 + 200. (This is an A.P, where a = 200 and d = 0) although this is rather trivial.

More interesting (and more profitable) is compound interest. The amounts form a G.P.

	After 1st year	**2nd Year**	**3rd Year**
10,000	10,000 x 1.02	10,000 x $(1.02)^2$	10,000 x $(1.02)^3$
	4th Year	**5th Year**	
	10,000 x $(1.02)^4$	10,000 x $(1.02)^5$	

Here a = 10,000, r = 1.02 $\left(\frac{102}{100}\right)$

After 5 years, the amount is Shs. 10,000 x 102^5 = Shs. 11,041

Both the sum of money and the rate of interest in this case are small, so the difference is not very great. Work out the difference for a sum of Shs. 100,000 at a rate of 10% for 5 years.

[Amount S.I, = Shs.150,000; amount C.I = Shs.161,051

$\therefore$ difference Shs. 11,051]

Also consider the example below:

Juma deposited Shs. 16,000 in bank which paid compound interest at the rate of 12% per annum. At the end of five years, he withdrew all his money. Using geometric series, find the amount of money he withdrew.

Solution:

Principal = Shs. 16,000

Interest for first year is 16,000 x $\frac{12}{100}$

Interest for 2^{nd} year is $16{,}000 \times \frac{112}{100} \times \frac{12}{100}$

Interest for 3^{rd} year is $16{,}000 \times \frac{112}{100} \times \frac{112}{100} \times \frac{12}{100}$

Interest for 4^{th} year is $16{,}000 \times \frac{112}{100} \times \frac{112}{100} \times \frac{112}{100} \times \frac{12}{100}$

Interest for 5 yer is $16{,}000 \times \frac{112}{100} \times \frac{112}{100} \times \frac{112}{100} \times \frac{112}{100} \times \frac{12}{100}$

Common ratio is $\frac{2^{nd} \text{ term}}{1^{st} \text{ term}}$

$$= \frac{6{,}000 \times \frac{112}{100} \times \frac{12}{100}}{6{,}000 \times \frac{12}{100}}$$

$$= \frac{112}{100}$$

$$= 1.12$$

1^{st} term $= 16{,}000 \times \frac{12}{100}$

Total interest $= \frac{a(r^n - 1)}{r - 1}$

$$= 16{,}000 \times \frac{12}{100} \frac{(1.12^5 - 1)}{012}$$

$$= 12197.5$$

Amount after five years

= Principal + total interest

= 16,000 + 12197.5

= Shs. 28197.5

ACTIVITY: MT/2/1-4

1. Prove that the series whose nth term is $(5n + 2)$ an A.P. Find the first term, the common difference and the sum of the first 10 terms.

2. Find the sum of the first 20 terms of the series $3 + 8 + 13 + 18$………

3. Prove that the series whose nth term is $4 \times 3n+1$ is a G.P. Find the first term, the common ratio and the sum to n terms.

4. Find how many terms of the G.P. $1 + 4 + 16$……are necessary for the sum to exceed 1000.

5. If for an A.P. the 7^{th} term is 13 and the sum of the first 14 terms is 203, find the 10^{th} term and the sum of the first 8 terms.

6. The sum of the first 10 terms of n A.P. is 120. The sum of the next 10 terms is 320. Calculate the first term and the common difference.

7. How many terms of the G.P., $2 + 6 + 18$……, must be taken to give a sum of 2186?

8. James deposited Shs. 20,000 in a fixed deposit account for a period of four years. The bank pays compound interest of 10% per year. If he withdrew all the money after four years, find how much he withdrew.

SUMMARY

In this unit, you have been introduced to a number of different sub-topics including:

1. Surds and Indices
2. Sequences and Series

GLOSSARY

TOPIC 1

- A SURD: Any root which has no exact value.

- AN IRRATIONAL NUMBER: any number which CANNOT be exposed in the form $\frac{p}{q}$ where p and q are whole numbers.

- TO RATIONALIZE: To remove the surd.

- The CHARACTERISTIC of a logarithm is determined by the power of ten, when the number is written in which the numbers lies e.g. a characteristic of 2 indicates that the number is between 10^2 and 10^3.

- The \MANTISSA' of a logarithm is the remainder of the logarithm the digits after the point which are determined from tables.

- An 'INDEX' (plural INDICES) is a power.

- The STANDARD FORM of a number is a x 10n where a is a number between one and ten and n is an integer.

- A LOGARITHM is a power of index.

TOPIC 2

- A SEQUENCE is a set of numbers, written in order which have some correction with each other.

- A SERIES is similar to a sequence but is written as the sum of the terms.

- AN ARITHMETIC PROGRESSION (A.P) is a sequence where the following term is found by adding (or subtracting) the same number each time. This number is called the 'COMMON DIFFERENCE'.

- A 'GEOMETRIC PROGRESSION' (G.P) is a sequence where the following term is found by multiplying (or dividing) by the same number each time. This number is called the 'COMMON RATIO'.

NOTES AND ANSWERS TO ACTIVITIES

ACTIVITY: MT/2/1-1

1. (i) $\sqrt{18} = 3\sqrt{2}$

   (ii) $\sqrt{50} = 5\sqrt{2}$

   (iii) $\sqrt{200} = 10\sqrt{2}$

   (iv) $\sqrt{7200} = 60\sqrt{2}$

2. (i) $\frac{2}{\sqrt{3}} = 2\frac{\sqrt{3}}{3}$

   (ii) $\frac{3}{\sqrt{5}} = \frac{3\sqrt{5}}{5}$

   (iii) $\frac{5}{\sqrt{7}} = \frac{5\sqrt{7}}{7}$

   (iv) $\frac{2}{\sqrt{11}} = \frac{2\sqrt{11}}{11}$

3. (i) $\sqrt{2} + 1$

   (ii) $\sqrt{3} + \sqrt{2}$

   (iii) $4 - \sqrt{15}$

4. $20\sqrt{2}$

5. $29 + \sqrt{5}$

6. 43

ACTIVITY: MT/2/1-2

1. (i) 2^8 (ii) 3^{10} (iii) 3^0 (iv) 6^3 (v) a^4

2. (i) $\frac{1}{16}$ (ii) $\frac{1}{16}$ (iii) 1 (iv) $\frac{1}{8}$ (v) $\frac{12}{13}$

3. 432

ACTIVITY: MT/2/1-3

1. 14800
2. 15.9
3. 2.11
4. 7.19
5. 4110

ACTIVITY: MT/2/1-4

1. $a = 7$, $d = 5$, sum of 10 terms = 295
2. Sum of 20 terms = 1010
3. $a = 36$, $r = 3$, $S_n = 18(3^n - 1)$
4. At least 6 terms.
5. 22, 44
6. 3, 2
7. 7
8. 29282

END OF UNIT ASSIGNMENT MT/2/1

Do this assignment in your exercise book or on a separate piece of paper. You are advised to read through the whole unit again before you do this assignment. Check your answers with those given at the end of the module.

1. Simplify the following if possible and rationalize surd denominations.

   (i) $\sqrt{243}$ (ii) $\frac{1}{\sqrt{8}}$ (iii) $\frac{(2 - 4\sqrt{2}}{(1 + 2\sqrt{2})}$

   (iv) $(3 - 5\sqrt{3})(3 + 5\sqrt{3})$ (v) $\frac{\sqrt{27+54}}{72-8}$

2. Simplify the following:

   (i) $(1024)^{2/5}$ (ii) 5^{-2} (iii) $\left(\frac{121}{169}\right)^{-3/2}$

   (iv) $\left(\frac{1}{2}\right)^{-1} + (32)^{4/5} - \left(\frac{4}{25}\right)^{-1/2}$ (v) $8^{-2/3}$

3. Simplify the following using logarithms, giving your answer correct to 3 significant figures:

   $$\sqrt{\frac{25.67x0.5981}{4203x0.01258}}$$

4. (i) Find the sum of 1 + 2 + 3 + ….. + 49
   (ii) Find the sum of 12, 5…… - 72
   (iii) Find the number of terms in each of the progressions in (i) and (ii).
   (iv) Find the sum of 3 + 9 + ….. as far as the 14^{th} term.
   (v) Given that a, b and 10 are in A.P.and their sum is -56, find a and b.

5. (i) Write down the 7^{th} term of 1, 5, 25………

   (ii) Write down the sum of 5 terms of the G.P. – 8, 24, -72,….(leave your answer in index form).
   (iii) Given that the third and fifth terms of a G.P. are 4 and 81 respectively, find the possible values of the common ratio.
   (iv) Given that the third and sixth terms of a G.P. are 2 and 54 respectively, find the first three terms and write down a formula for S_n.

LEARNING OUTCOMES – SELF CHECKING EXERCISE

You have completed Unit 1 of Module Mt/2. The learning outcomes are listed below: Place a tick in the column that best reflects your learning.

	LEARNING OUTCOMES	***NOT SURE***	***SATISFACTORY***
1.	I can manipulate surds and simplify expression with surds in them.		
2.	I can use the laws of indices to evaluate and simplify expression involving indices.		
3.	I can use logarithm tables in computation involving manipulation, division, roots and powers.		
4.	I can identify an Arithmetic progression, its common difference, the n^{th} term and find the sum of n terms.		
5.	I can identify a Geometric progression, its common ratio, the n^{th} term and find the sum of n terms.		

ANSWERS TO END OF UNIT ASSIGNMENTS MT/2-1

1. (i) $9\sqrt{3}$ (ii) $\frac{\sqrt{2}}{4}$ (iii) $\frac{2(4\sqrt{2}-9)}{7}$

   (iv) -66 (v) $\frac{9}{8}$

2. (i) 16 (ii) 0.04 (iii) $\frac{2197}{1331}$ (iv) $15\frac{1}{2}$ (v) $\frac{1}{4}$

3. 0.539

4. (i) 1225 (ii) -390 (iii) 49 terms and 13 terms (iv) 588

   (v) $a = -47\frac{1}{3}$ $b = -18\frac{2}{3}$

5. (i) 5^6 (ii) $-2(1 + 3^5)$ (iii) $r = \pm\frac{9}{2}$

   (iv) First 3 terms: $\frac{2}{9}$, $\frac{2}{3}$, 2. $S_n = \frac{(3^n - 1)}{9}$

UNIT 2

STATISTICS

MT/2/2

INTRODUCTION

Dear Student,

You are welcome to Unit 2 of Module 2 which is intended to introduce you to how to deal with information in a numerical form which is called data. The emphasis will be more to grouped data.

AIMS

The aims of this unit are to enable you:

1. Develop manipulative skills required for handling data.

2. Understand statistics as a discipline of mathematics and use this knowledge to handle the Primary Syllabus, especially the section on statistics effectively.

3. Use statistics in your every day life.

4. Build a background to an academic understanding of statistics and its application in research at a later stage.

5. Increase your knowledge of measures that are used in statistics.

6. Apply different methods in solving measures of dispersion.

SPECIFIC OBJECTIVES

By the end of this unit, you should be able to:

1. Group raw data

2. Tabulate data by grouping it.

3. Draw histograms and cumulative frequency curves (ogives) using the tabulated data.

4. Draw frequency polygon using tabulated data.

5. Calculate the mean of the tabulated data using assumed mean and the method of coding.

6. Find the mode using the histogram.

7. Compute the median using the cumulative frequency curve (ogive).

8. Find the mode and median for grouped data using formulate.

9. Name the different measures of dispersion that are used in statistics.

10. Evaluate the different measures of dispersion for a given set of data using a variety of methods.

11. Apply the knowledge of quartiles and percentiles to solve statistical problems.

UNIT ORGANIZAITON

This unit is organized into 6 topics:

Topic 1: The frequency distribution of grouped data; has four sub-topics:

(a) Grouped data
(b) Frequency table of grouped data.
(c) Class limits
(d) Class boundaries

Topic 2: Graphical Representation of grouped data; has four sub-topics:

(a) Histograms
(b) Frequency polygons
(c) Cumulative frequency curves (ogives)

Topic 3: The Measure of Central tendency of grouped data; has three sub-topics:

(a) Mean
(b) Mode
(c) Median

Topic 4: Standard Deviation and Variance; has 3 sub-topics

(a) Calculate standard deviation and variance from raw data.
(b) Calculating standard deviation and variance from frequency distribution.
(c) Use of coding method to find the standard deviation.

Topic 5: Quartiles, Percentiles and Range; has 3 sub-topics

(a) Methods of calculating quartiles and percentiles.
(b) Approximating quartiles and percentiles from cumulative frequency curve (ogive).
(c) Approximation of range and inter quartile range form cumulative frequency curve.

Topic 6: Application of Standard Deviations

TOPIC 1: FREQUENCY DISTIRBUTION OF GIVEN DATA

In this topic you will learn about:

- Grouped data
- Frequency table of grouped data
- Class limits
- Class boundaries

(a) **GROUPED DATA**

Raw data presented in different groups or classes is what we call grouped data. Normally we choose from six to twelve classes unless the order of writing these classes is given. A better choice of class size (interval) is very important so as to cover the range of the raw data.

Example:

The following marks out of 100 were obtained by twenty eight DEPE students in a test on 'Statistics'.

Table 1:

68	80	60	66	86	82
48	94	48	46	66	78
46	60	66	48	46	68
72	42	46	86	84	66

Write all classes starting from 40 – 49, 50 – 59, etc, which will include all the data given.

Solution:

<u>Classes</u>

40 - 49
50 - 59
60 - 69
70 - 79
80 - 89
90 - 99

☞ **Note:** All classes should be of the same class size (class interval). The above classes have a class interval of 10.

Class interval is the difference of the lowest value of a class from the lowest value of the next class.

ACTIVITY: MT/2/2-1

Use table 1 to answer the following questions:

1. what is the highest mark?

2. What is the lowest mark?

3. Find the difference between the highest mark and the lowest mark. This value is called the range.

4. Using a class interval of 5, write down the different classes for the data.

5. Why you think we do not use classes below 40 marks and those above 100 marks for the given data?

6. What advantage could grouped data have over ungrouped data for many values in the raw data?

✓ Check your answers with those given at the end of this unit.

(i) Guidelines to choice of Class Interval

- Find the range of the data.
- Decide on the number of classes (from 6 to 12) you wish to use.
- Divide the range by the number of classes chosen. The result is an approximation of the class size.

Example:

Using data in Table 1:

- The range is: 94 - 52 = 52
- For 6 classes, the class interval will be $\frac{52}{6} = 8.6 = 9$

- The classes may then be:

  42 - 50
  51 - 59
  60 - 68
  69 - 77
  78 - 86
  87 - 95

☞ **Note**: For convenience, class sizes of 10 and 5 are often used.

ACTIVITY: MT/2/2-2

The following marks were scored by 40 pupils in a mathematics test.

Table 2:

54	42	73	54	58	85	52
60	58	48	70	52	53	53
45	60	55	50	53	75	58
63	58	57	82	30	35	49
48	53	52	57	60	25	65
72	54	55	68	28		

1. Use the marks to:

   (i) Write down the highest mark.
   (ii) Write down the lowest mark.
   (iii) Find the range of the distribution.

2. Using a class interval of five:

   (i) Group the data in table 2.
   (ii) What is the last class?

✓ Check your answers with those given at the end of this unit.

(b) FREQUENCY TABLE OF GROUPED DATA

When dealing with raw data with many values, we can break up the values and distribution then into classes. The number of members falling in a class is the class frequency. A class with the highest frequency is called the modal class.

Example:

The following are marks obtained by 45 students in a Mathematics Examination at Kibuli TTC (DEPE centre).

Table 3:

72	76	82	76	59	67	73	80	90
63	78	73	91	66	88	63	74	86
67	71	51	86	82	76	48	72	87
72	86	76	49	96	71	72	85	87
71	83	66	71	51	46	81	82	64

Make a frequency table for the above data.

Solution: Using a class interval of 10, we get Table 4.

Table 4:

Classes	Tally	Frequency (f)
40 - 49	///	3
50 - 59	//	3
60 - 69	~~////~~ /	7
70 - 79	~~////~~ ~~////~~ ~~////~~ /	16
80 - 89	/~~///~~ ~~////~~ ///	13
90 - 99	///	3

Note: The above frequency table has a tally column.

To get the tallies, for any members you identify for a particular class you attach a stroke (tally).

Consequently, you will form a tally column. The maximum number of tallies in a bundle is five. The fifth tally ties the bundle as indicated in table 4. compare this with the five fingers on each hand and the five toes on each foot.

The number of tallies corresponding to a class corresponds to the frequency.

☞ **Caution:** When presenting the frequency table the tally column should be there. The most frequent class (medal class) is 70 – 79.

***A*CTIVITY: MT/2/2-3**

1. The number of tomatoes collected daily (except on Sundays) over a period of six weeks by a farmer were as follows:

75	79	92	83	87
97	79	87	89	78
84	99	92	97	99
89	80	85	87	91
100	80	89	80	94
77	97	91	82	88
94	87	95	100	95
94				

(a) Construct a grouped frequency table with class interval of 5 starting at 75.
(b) Write down the modal class.

2. The following are the length in centimeters of timber.

353	358	345	301	317
337	300	299	316	308
299	323	312	318	324
345	350	341	318	332
337	350	300	308	345
306	352	331	358	340
300	341	345	317	345
346				

Make a frequency table for grouped data of the above information using:

(i) Class interval of 10 cm.
(ii) Find the modal class.

✓ Check your answers with those at the end of this unit.

(c) CLASS LIMITS

the top and bottom values of the classes are the class limits.

In the example in Table 1:

40 and 49 are class limits for the class 40 - 49.
40 is the lower class limit and 49 is the upper class limit.

***A*CTIVITY: MT/2/2-4**

Write down class limits for the different classes in table 1.

Note: The average of these two class limits is called the class mark or midpoint.

✓ Check your answers with those at the end of this unit.

(d) CLASS BOUNDARIES

The class boundaries are the numbers that fall halfway between the upper class limit of one class and the lower class limit of the next class.

Example:

Referring to the example in table 1, the classes were as follows:

40 - 49
50 - 59
60 - 69
70 - 79
80 - 89
90 - 99

We shall find the class boundaries for the first class, 40 - 49.

Solution:

Using the same interval, use a class just before the lower class limit for the class 40 - 49 which is 30 - 39,

Then the lower class boundary is; $\frac{39 + 40}{2} = 39.5$

Similarly, the upper class boundary is; $\frac{49 + 50}{2} = 49.5$

Caution: Some references do not bother with these class boundaries and only write about class limits to mean class boundaries.

Please, watch out!

A**CTIVITY: MT/2/2-5**

1. Find the lower and upper class boundaries for the classes.

   (i) 70 - 79
   (ii) 90 - 99

2. Find the averages of the class boundaries for each class in No. 1

3. Find the averages of the class limits for each class in NO.1

4. Compare the results in NOs. 2 and 3.

5. Write a simple formula for the class mark of a given class.

✓ Check your answers with those at the end of this unit.

N.B: ***The difference between the upper class boundary and the lower class boundary gives the class interval/size/width.***

TOPIC 2: GRAPHICAL REPRESENTATION OF GROUPED DATA

In Topic 1, we studied frequency distributions of grouped data. In this topic we are to study the various graphical methods of representing grouped data from a frequency table.

(a) HISTOGRAM

A histogram is one way by which a frequency distribution can be represented. It is a bar chart in which the areas of the rectangles are proportional to the frequencies. The bars are of the same width and attached to one another. Here, you need to know how to find class boundaries.

Example:

Using the frequency table for grouped data, in Activity MT/2/2-2, NO. 2 (i), draw a histogram.

Solution:

<u>Classes</u>	<u>Class boundaries</u>	<u>Frequency</u>
25 - 29	24.5 - 29.5	2
30 - 34	29.5 - 34.5	1
35 - 39	34.5 - 39.5	0
40 - 44	39.5 - 44.5	1
45 - 49	44.5 - 49.5	4
50 - 54	49.5 - 54,5	11
55 - 59	54.5 - 59,5	9
60 - 64	59.5 - 64.5	4
65 - 69	64.5 - 69.5	2
70 - 74	69.5 - 74.5	3
75 - 79	74.5 - 79.5	1
80 - 84	79.5 - 84.5	1
85 - 89	84.5 - 89.5	1

To draw a histogram, frequencies are marked on the y - axis. To find the scale to be used on the y – axis, consider the highest and lowest frequencies.

In this example, the highest frequency is 11. The scale can be 1 cm to 1 unit of frequency on the y – axis and 1 cm as the width of the bar on the x – axis.

*

☞ **Note**: Since the first class does not start at zero, leave a gap then draw the first bar.

ACTIVITY: MT/2/2-6

Below are marks obtained by fifty pupils of Primary Seven in Mathematics.

35	54	50	59	50	61	52	59
38	55	65	57	54	70	58	69
40	63	50	59	60	50	67	64
42	57	59	62	63	51	58	60
48	36	45	47	49	50	57	61
60	50	58	56	71	67	62	60
65	49						

1. Write down the (i) highest score
   (ii) lowest score
2. Using a class interval of 5, construct a frequency table for the data and draw a histogram.

✓ Check your answers with those at the end of this unit.

(b) **FREQUENCY POLYGON**

This is obtained by joining the midpoints of the bars of the histogram of the different classes using line segments.

Example:

The number of accidents that took place in Kampala in July 1981 was recorded as follows:

No of accidents	2 - 4	5 - 7	8 - 10	11 - 13	14 - 18
No. of days	2	5	10	8	5

Draw a frequency polygon representing data.

Solution:

No. of accidents	Mid-points of class (x)	Frequency
2 - 4	3	2
5 - 7	6	5
8 - 10	9	10
11 - 13	12	8
14 - 18	16	5

Frequency Polygon

*

☞ **Note**: When you join the ends of the diagram above to the x – axis, then you will have a polygon, hence frequency polygon.

A**CTIVITY: MT/2/2-7**

Draw a frequency polygon for the following data of speeds of cars traveling on a highway.

Speed km/h	31-35	36-40	41-45	46-50	51-55	56-60	61-65	66-70	71-75	76-80
No. of cars	2	6	5	7	25	49	33	8	3	4

✓ Check your answers with those at the end of this unit.

(c) CUMULATIVE FREQUENCY CURVE (OGIVE)

Before drawing a cumulative frequency curve, let us consider another way of showing data. This is through developing a cumulative frequency table. The cumulative frequency table is formed by adding frequencies and writing a total for each frequency added.

To make this clear, let use the example below:

Example:

Construct a cumulative frequency table from the frequency distribution below:

The table shows the heights of 100 clonal coffee seedlings after 3 months in a nursery bed.

Height (cm)	**Number of Clonal coffee seedlings (Frequency)**
60 - 62	5
63 - 65	18
66 - 68	42
69 - 71	27
72 - 74	8

Remember that the cumulative frequency is a total obtained after adding a given frequency to the previous total. You should note that the members of each class are less than the upper class boundaries of the class. For example:

57 - 59 means less than 59.5 and 60 – 62 means less than 62.5

From table 12, we can now write down a cumulative frequency table (for coffee seedlings. Here we use the upper class boundary for each class.

Height (up to)	**Cumulative frequency**
59.5	0
62.5	5
65.5	23
68.5	65
71.5	92
74.5	100

☞ **Note:** 59.5 is the upper class boundary for the values from 0 to 59. Using the table above, you notice that the final total under the cumulative frequency is the total number of coffee seedling. This is the total frequency.

Now, we are ready to draw a cumulative frequency curve. This is also referred to as the ogive.

Observe the cumulative frequencies and decide on the scale for the y-axis. Count the upper class boundaries then decide on the scale to be used on the x-axis. The points should be joined using a smooth curve.

Cumulative Frequency Curve

*

***A*CTIVITY: MT/2/2-8**

The following are examination marks for 15 students.

84	75	91
61	75	67
72	87	79
75	79	83
77	51	69

1. Construct a grouped frequency distribution table starting with 50 – 59, 60……..69,……..

2. Make a cumulative frequency distribution table.

3. Draw a cumulative frequency curve.

✓ Check your answers with those at the end of this unit.

TOPIC 3: MEASURES OF CENTRAL TENDENCY OF GROUPED DATA

In Topic 2, we studied the various graphical methods of representing grouped data using a frequency table.

In this topic we shall discuss three measures of central tendency namely; mean, mode and median.

(a) THE MEAN

(i) Calculating mean using assumed mean

This method of calculating mean reduces considerably the amount of calculation involved. For this, we choose an estimated mean which is called a working mean or an assumed mean usually denoted as A. We find the differences (deviation) of the assumed mean from each score and then find the mean of these deviations. Adding the mean deviation to the assumed mean, we get the actual mean.

Example:

The table shows the results of 25 nursery children who wanted to joint Primary One.

Marks	**No. of children (f)**
50 - 59	2
60 - 69	4
70 - 79	11
80 - 89	6
90 - 99	2

Estimate the mean using an assumed mean.

Solution:

Le the assumed mean be A = 74.5

Marks	Class Marks (x)	Frequency (f)	x - A = d	fd
50 - 59	54.5	2	20	$^{-}40$
60 - 69	64.5	4	10	$^{-}40$
70 - 79	74.5	11	0	0
80 - 89	84.5	6	10	60
90 - 99	94.5	2	20	40
		$\sum f = 25$		$\sum fd = 20$

Hint: Assume a value that is midway of all the class marks. Suppose 74.5 is the assumed mean (A). The class mark of the middle class is usually taken as the suitable assumed mean.

The actual mean $x = A + \frac{\Sigma fd}{\Sigma f}$

$$= 74.5 + \frac{20}{25}$$

$$= 74.5 + 0.8$$

$$= 75.3$$

***A*CTIVITY: MT/2/2-9**

In a class of 40 students, the teacher needed to estimate the mean age of these students.

Age	Frequency
13 - 15	6
16 - 18	8
19 - 21	20
22 - 24	4
25 - 27	2

Using 17 as the assumed mean, calculate the mean age.

✓ Check your answers with those at the end of this unit.

(ii) Calculating the Mean using the Coding Method

In the example below, we shall find the mean using a coding method. This method is similar to that of finding the mean using an assumed mean.

Example:

Use coding method to find the mean of the data below:

Class	Frequency (f)
6 - 10	8
11 - 15	10
16 - 20	12
21 - 25	20
26 - 30	16
31 - 35	11
36 - 40	3

Solution:

Class	Class marks	Coding Numbers (u)	f	(f x u)
6 – 10	8	$^-3$	8	$^-24$
11 – 15	13	$^-2$	10	$^-20$
16 – 20	18	$^-1$	12	$^-12$
21 – 25	23	0	20	0
26 – 30	28	1	16	16
31 – 35	33	2	11	22
36 – 40	38	3	$\Sigma f = 80$	$\Sigma fu = {}^-9$

$$\text{Mean} = \text{The class mark} + \frac{\Sigma fu}{\Sigma} \times \text{class interval}$$

$$= 23 + \frac{^-9 \times 5}{80}$$

$$= 23 - \frac{9}{16}$$

$$= 22\frac{7}{16}$$

***A*CTIVITY: MT/2/2- 10**

Given the table below, use coding method to find the mean.

Classes	**Frequency (f)**
110 – 114	5
115 – 119	6
120 – 124	17
125 – 129	21
130 – 134	19
135 – 139	16
140 – 144	12
145 – 149	4

✓ Check your answers with those given at the end of this unit.

(b) THE MODE

(i) Finding the Mode from a Histogram

Histogram (Frequency Distribution Diagram)

Histogram (Frequency Distribution Diagram)

*

When finding the mode from a histogram, draw a line segment from A to B then from C to D. Where the two lines segments meet, draw a perpendicular line to the x – axis. Let the perpendicular line meet the x – axis at a point E.

Locate this point E by use of your scale. From the graph, the point is given by 49.5 + 4.0 = 53.5, hence the modal mark is 53.5.

(ii) Finding the mode using the formula

$$\text{Mode} = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)C$$

Where: L_1 is lower class boundary of the modal class.

Δ_1 = frequency of the modal class – frequency of the class just before the modal class.

Δ_2 = frequency of the modal class – frequency of the class just after the modal class.

C is the class interval.

Using the previous example;

$$\Rightarrow \quad \Delta_1 = 11 - 4 = 7$$
$$\Delta_2 = 11 - 9 = 2$$
$$C = 5$$
$$L_1 = 49.5$$

Then the mode $= 49.5 + \frac{7 \times 5}{7 + 2}$

$$= 49.5 + 3.9$$
$$= 53.4$$

Use either the histogram or the table (arranged in order) to determine the unknown values Δ_1 and Δ_2

ACTIVITY: MT/2/2-11

The distribution of a group of people is given in the table below:

Age	15 - 19	20 - 24	25 - 29	30 - 34	35 - 39	40 - 44
Frequency	22	35	13	9	24	30

Age	45 - 49	55 - 59	60 - 64	65 - 67	70 - 74	75 - 79
Frequency	40	22	35	13	9	24

1. Draw the histogram for this distribution.
2. (i) Find the mode from the histogram.
   (ii) Use the formula to find the mode.

✓ Check your answers with those given at the end of this unit.

(c) THE MEDIAN

(i) Finding the Median using the Cumulative Frequency Curve (the ogive)

Take the last value of the cumulative frequency (N), which is the total frequency. Divide this value by 2 and draw a line across the graph from the cumulative frequency $= \frac{N}{2}$ until it cuts the curve and drop a vertical line to the horizontal axis.

Example:

Find the median for the data given below:

Class	**Frequency**
0 – 9	2
10 – 19	14
20 – 29	24
30 – 39	12
40 – 49	8

Solution

Cumulative frequency distribution

Class	Upper boundary	Frequency (f)	Cumulative Frequency (cf)
0 – 9	9.5	2	2
10 – 19	19.5	14	16
20 – 29	29.5	24	40
30 – 39	39.5	12	52
40 – 49	49.5	8	60

Solution Cumulative Frequency curve (ogive)

*

The last value for cumulative frequency is N = 60

$$\frac{N}{2} = \frac{60}{2} = 30$$

The median is 19.5 + 6.5 = 26, corresponding to the cumulative frequency of 30.

(iii) Finding the Median using the formula

$$\text{Median} = L + \left(\frac{\frac{N}{2} - B)C}{f}\right.$$

Where:

L is the lower class boundary for the median class.
N is the total frequency (the last value for the cumulative frequency).
B is the cumulative frequency for the class just before the median class.
f is the frequency of the median class.

$$\therefore \quad \text{Median} = 19.5 + \frac{(30 - 16)\,10}{24}$$

$$= 19.5 + \frac{14 \times 10}{24}$$

$$= 19.5 + \frac{140}{24}$$

$$= 19.5 + 5.83$$

$$= 25.33$$

ACTIVITY: MT/2/2-12

The age distribution of a graph of people is given in the table below:

Age	15 - 19	20 - 24	25 - 29	30 - 34	35 - 39	40 - 44	45 - 49
Frequency	22	35	13	9	24	30	40

Age	50 - 54	55 - 59	60 - 64	65 - 69	70 - 74	75 - 79
Frequency	19	15	10	9	9	5

1. Draw a cumulative frequency curve (ogive) for the data.
2. (i) Find the median using the ogive.
   (ii) Use the formula to find the median.

✓ Check your answers with those given at the end of this unit.

TOPIC 4: STANDARD DEVIATION AND VARIANCE

In this topic, you will learn about:

- Calculating the standard deviation and variance from raw data.
- Calculating the standard deviation and variance from frequency distributions.
- Use of the coding method to find the standard deviation.

(a) CALCULATING THE STANDARD DEVIATION AND VARIANCE FORM RAW DATA

(i) Introduction

In Module, MT/ and Module MT/2, topic 3, you learnt about the different measures of location (or central tendency). These were given as the mean, median and mode. Please revise these measures so that you are very familiar with them and how they are calculated before proceeding with this topic.

We are now goring to learn about the measures of dispersion or variation of a given set of data. These measures tell us how the observations are spread, particularly how they are spread about the middle value. Let us illustrate this with a simple example. Consider the two sets of data below:

(a)	11	12	13
(b)	1	12	23

What is the mean for the observations in (a)?
What is the mean for the observations in (b)?
What can you say about the means for the two sets of data?

Right! The mean of each set is 12. Now look at each set critically and then compare the values given in each set. You must have noticed that there is clearly more dispersion (variation or spread) in the second set. We shall therefore proceed to get some measures of dispersion.

Initially, we may be tempted to find out how each individual observation varies from the mean. We may therefore find the values x_i - x for each set. Let us proceed as follows:

Set	(i)	x_i	11	12	13
		$x_i - \bar{x}$	-1	0	1

Remember, the mean for the observation is 12. Now since we are interested in the whole set of data, we could think of finding the average of the values x_i – x, which are called the deviations from the mean.

Using our knowledge of calculating the mean, let us find the mean of ¯1, 0, 1 since these are the actual values of x_i - x.

That is, their mean $= \sum\left(\frac{x_i - \bar{x}}{n}\right)$

Where n = number of items

i.e. mean $= \frac{-1+0+1}{3}$

$= \frac{0}{3}$

$= 0$

What if we use set (ii)? We shall have:

x_i	1	12	23
$x_i - \bar{x}$	¯11	0	11

and then $\sum\left(\frac{x_i - \bar{x}}{n}\right)$

$= \frac{{}^{-}11 + 0 + 11}{3}$

$= \frac{0}{3}$

$= 0$

In both cases, we see that the average of the deviations from the mean is zero. Indeed, in all cases, it will always be zero. Can you try to think of why? Consequently, it is not a useful measure of the dispersion about the mean.

You may have noticed in the two cases above that, the negative deviations from the mean are equal to the positive deviations. This is why we end up with a sum of zero! The measures of dispersion that we are about to look at avoid getting a sum of zero by squaring the deviations from the mean.

(ii) **The Standard Deviation from Raw Data**

The standard deviation is a very important measure of dispersion and is calculated through the following steps:

Step 1: Find the data mean $\bar{x}$

Step 2: Determine the n values $x_i - \bar{x}$

Step 3: Square these deviations to obtain $(x_i - \bar{x})^2$

Step 4: Add all the squared deviations to obtain $\sum (x_i - \bar{x})^2$

Step 5: Divide by n to get the mean of the values in Step 4: i.e $\sum \left(\frac{x_i - \bar{x}}{n} \right)^2$

Step 6: Take the square root of the result in Step 5 to end up with

$$\sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

What you have now is referred to as the standard deviation (s d) and is commonly denoted by s.

Hence, s $= \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$

It is worth noting a few things about the steps involved in getting the formula standard deviation:

The values x_i - x in step 2 add up to zero always. We try to avoid this by squaring in step 3.

We make an attempt of finally getting the average deviations from the mean that are free from the effect of squaring in Step 3 by finding the square root in Step 6.

Example 1:

On six consecutive days, Mary received 9, 7, 11, 10, 13 and 7 service calls. Calculate the standard deviation for the number of calls that Mary received.

Solution:

Remember that sd, measures dispersion about the mean. The firs step therefore is to find the mean.

Mean x = $\underline{9 + 7 + 1 + 1 + 10 + 13 + 7}$

6

$$= \frac{57}{6}$$

$$= 9.5$$

We can now organize our work in tabular form to work out steps 2 and 3 as follows:

x_i	$x_i - \bar{x}$	$(x - \bar{x})^2$
9	-0.5	0.25
7	-2.5	6.25
11	1.5	2.25
10	0.5	0.25
13	3.5	12.25
7	-2.5	6.25
	$\sum(x_i - \bar{x}) = 0$	$\sum(x_i - \bar{x})^2 = 27.50$

Before we proceed, do you notice that, as earlier pointed out $\Sigma(x - \bar{x}) = 0$? Good. Now we shall quote the formula for sd and then substitute; i.e.

$$s = \sqrt{\frac{\sum(xi - \bar{x})2}{n}}$$

$$= \sqrt{\frac{27.50}{6}}$$

$$= \sqrt{4.58}$$

$$= 2.14$$

$$= 2.14 \text{ (2 d.p)}$$

Example 2:

When asked the distances of their homes in km from school, 5 pupils had the following answers.

80, 3, 90, 272, 80

Find the standard deviation for these distances.

Solution:

$$\bar{x} = \frac{80 + 3 + 90 + 272 + 80}{5}$$

$$= \frac{525}{5}$$

$= 105$

x	$x_i - x$	$(x_i - x)^2$
80	$^-25$	625
3	$^-102$	10404
90	$^-15$	225
272	167	27889
80	$^-25$	625
	$\sum (x_i - x) = 0$	$\sum (x_i - x)^2 = 3968$

Again we note that $\sum (x_i - \overline{x}) = 0$

In future, we shall not be computing this sum.

Now we have s $= \sqrt{\frac{(xi - x)2}{n}}$

$$\sqrt{\frac{39768}{5}}$$

$$= \sqrt{7953.6}$$

$$= 89.2 \text{ (1 d.p)}$$

***A*CTIVITY: MT/2/2-13**

1. When some parents were asked the ages in years of their children, they gave the following answers: 8, 3, 5, 7, 11, 9, 7, 6, 7, 5, 8, 1. Find the standard deviation for these observations.

2. At a certain polling station, the ages of the registered voters were recorded as: 28, 18, 32, 27, 42, 23, 49, 28, 33, 30. What is the standard deviation of these ages?

3. When asked about the age of the oldest living member of their district officials reported the following observations: 76, 79, 85, 86, 76, 75, 86, 105, 88, 104, 102, 103, 89, 92, 95. Calculate the standard deviation to 1 decimal point.

4. Mr. Kintu gave a test to his class and marked it out of 50. Pupils had the scores below:

3	4	2	24	28
3	13	3	8	13
5	3	4	25	29
3	21	19	20	22
20	21	20	22	21

Calculate the standard deviation of the scores.

✓ Check your answers with those given at the end of this unit.

(ii) The Computing Formula for Standard Deviaiton

We have used the formula

$$s = \sqrt{\frac{\sum(xi - x)2}{n}}$$ to obtain the standard deviation

Squaring both sides of the formula gives us:

$$s^2 = \frac{\sum (x_i - \overline{x})^2}{n}$$

$$= \frac{\sum (x_i - \overline{x})^2}{n}$$ if we drop the subscription

Expanding $(x - \overline{x})^2$ we get: $\sum(x - \overline{x})^2$

$$= \quad \sum (x^2 - 2xxx + (x)^2$$

So that $s^2 = \dfrac{\sum (x^2 - 2\overline{x}x + (\overline{x})^2}{n}$

Opening the bracket, we shall have: $s^2 = \dfrac{\sum x^2 - \sum 2xx + \sum (x)^2}{n}$

Note that x is a constant so we can write it outside the summation sign. You should also recall that the limits of our summation vary from = 1 to = n (See Module MT/1).

We then have $s^2 = \dfrac{\sum x^2}{n} - \dfrac{2\overline{x\sum x}}{n} + \dfrac{(n\overline{x})^2}{n}$

$$= \sum \frac{x^2}{n} - 2\overline{\overline{x}x} + (\overline{x)^2}$$

Since $\dfrac{\sum x}{n} = \overline{x}$ and $\sum_{i=1}^{n} 1 = n$, in other words, here we add 1 to itself, n times.

$\therefore \quad s^2 = \dfrac{\sum x^2}{n} - (\overline{x)^2}$

Taking square roots of both sides, we end up with:

$$s = \sqrt{\frac{\sum x^2}{n} - (\overline{x)^2}}$$

This is known as the computing formula for the standard deviation. It can also be written in the form:

$$s = \sqrt{\frac{x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

This last form works with the raw scores of x and avoids using a rounded value of x. It therefore yields a more accurate value of s. In most cases, it also makes the working less cumbersome. We shall now work through the two examples we had earlier on using this last form of the computing formula.

Example 1:

Use the computing formula to find the standard deviation of the number of calls May received on six different days.

Solution:

The calls were recorded as 9, 7, 11, 10, 13 and 7. In tabular form we have:

x	x^2
7	49
7	49
9	81
10	100
11	121
13	169
$\Sigma x = 57$	$\Sigma x^2 = 569$

Thus $$s = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$= \sqrt{\frac{569}{6} - \frac{3249}{36}}$$

$$= \sqrt{\frac{3414 - 3249}{36}}$$

$$= \sqrt{\frac{165}{36}}$$

$$= \sqrt{4.5833333}$$

$$= 2.14 \text{ (2 d.p)}$$

Example 2:

Use the computing formula for the standard deviation to find s for the distances in km of the homes of the 5 pupils from school.

Solution: The distances are 80, 90, 272, 80

In table form we have:

x	x^2
3	9
80	640
80	640
90	810
272	73984
$\Sigma x = 525$	$\Sigma x^2 = 94893$

Then s $= \sqrt{\frac{x^2}{n} - \left(\frac{x}{n}\right)^2}$

$= \sqrt{\frac{94893}{5} - \left(\frac{x}{n}\right)^2}$

$= \sqrt{\frac{94893}{5} - \frac{275625}{25}}$

$= \sqrt{\frac{474465 - 275625}{25}}$

$= \sqrt{\frac{198840}{25}}$

$= \sqrt{7953.6}$

$=$ 89.2 (1 d.p)

As you may have noticed, the formulae are equivalent and gives us the same values for the standard deviation.

***ACTIVITY:* MT/2/2-14**

Work through all the examples of Activity MT/2/2-1 using the computing formula. Compare your answers with those given for MT/2/2-13

✓ Check your answers with those given at the end of this unit.

(iii) The Variance from Raw Data

The variance is the square of the standard deviation and it is also a useful measure of dispersion. However, variance is not as commonly used as standard deviation. Since we have denoted sd with s variance will be denoted as s^2. Then using the formulae we have derived:

(a) from s $= \sqrt{\frac{\sum\left(x_1-(\bar{x})\right)^2}{}}$

$$s^2 = \frac{\sum\left(x_1-(\bar{x})\right)^2}{}$$ gives the variance for the observations x_i

(b) from s $\sqrt{\frac{\sum x^2}{n}-\left(\frac{\sum x}{n}\right)^2}$

We have $s^2 = \frac{\sum x^2}{n}-\left(\frac{\sum x}{n}\right)^2$ as the computing formula for the variance.

In short, we square the standard deviation to get the variance. Alternatively, as we work towards getting the standard deviation, before taking the square root, we have a variance.

We therefore note the following:

1. The variance of the calls Mary receives on 6 different days is given by:
   $s^2 = 4.58$ (2d.p)

2. The variance of the pupils' home distances from school is given by:
   $s^2 = 7953.6$ (1 d.p).

ACTIVITY: MT/2/2-15

Write down the variance in each for the questions in Activity: MT/3/2-1

✓ Compare your answers with those given at the end of this unit.

(b) CALCULATING THE STANDARD DEVIATION AND VARIANCE FROM FREQUENCY DISTRIBUTIONS

As you are already aware, large numbers of data can be more easily handled with the use of frequency distributions. You will also recall that if the observations are ungrouped, each item is multiplied by its corresponding frequency. On the other hand, if observations are grouped, the midpoints are multiplied by their respective group frequencies. We are now going to introduce the idea of frequency in our calculations of standard deviation and variance.

(i) Ungrouped frequency distributions

From the formula s $= \sqrt{\dfrac{\left(x_1 - (\overline{x}\right)^2}{}}$

We realize that for a frequency distribution, each deviation from the mean (x_i - x) has a corresponding frequency f_i. We therefore sum the values f_i (x_i – x)^2 so that our formula becomes:

$$s = \sqrt{\frac{f_1\left(x_1 - (\overline{x}\right)^2}{}}$$

We shall illustrate this using an example before looking at the equivalent computing formula.

Example

The grade point averages (GPA) of 15 colleges graduates are given in the table below: calculate the standard deviation. What is the variance of the GPA?

Grade point	1	2	2	4	5	6
Frequency	1	2	3	5	3	1

Solution

In tabular form we have:

x_i	f_i	$f_i x_i$	$\left(x_1 - (\overline{x})\right)$	$\left(x_1 - (\overline{x})\right)^2$	$f_1\left(x_1 - (\overline{x}\right)^2$
1	1	1	-2.7	7.29	7.29
2	2	4	-1.7	2.89	5.78
3	3	9	-0.7	0.40	1.47
4	5	20	0.3	0.09	0.45
5	3	15	2.3	1.69	5.07
6	1	6	2.3	5.29	5.29

$\sum f = 15 \quad \sum f_1 = 55\,\Sigma \qquad\qquad \sum f_1\left(x_1 - (\bar{x})\right)^2 = 25.35$

Then x = $\dfrac{f_i x_i}{\sum f}$

$= \dfrac{55}{15}$

$= 3.7$ (1 d.p)

and s $= \sqrt{\dfrac{f_1\left(x_1 - (\bar{x})\right)^2}{n}}$

$= \sqrt{\dfrac{25.35}{15}}$

(Note that n = $\sum f$

$= \sqrt{1.69}$

$= 1.3$

From this example, you should have noticed how cumbersome it is to work out all the column in the table.

How heavy could the working get if you had bigger x values (say 100,110….) and higher frequency values? What if there are 10 or more x – value? To lessen the burden of working and to avoid errors, we shall use the computing formula.

The equivalent formula for s = $\sqrt{\dfrac{\sum f_1\left(x_1 - (\bar{x})\right)^2}{n}}$

is s = $\sqrt{\dfrac{\sum f_1 x_1^2}{} - \left(\dfrac{\sum f_1 x_1}{n}\right)^2}$

Compare with the formulae we had for raw data! Let us now use this formula in an example. The table will now look like this:

x	f	fx	x2	fx2
1	1	1	1	1
2	2	4	4	8
3	3	9	9	27
4	5	20	16	80
5	3	15	25	75

6	1	6	36	36
	$\sum f = 15$	$\sum fx = 55$		$\sum fx^2 = 227$

Then s $= \sqrt{\dfrac{\sum fx^2}{n} - \left(\dfrac{\sum fx}{n}\right)^2}$

$$= \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

$$= \sqrt{\frac{227}{15} - \left(\frac{55}{15}\right)^2}$$

$$= \sqrt{\frac{227}{15} - \frac{3025}{225}}$$

$$= \sqrt{\frac{380}{225}}$$

$$= \sqrt{1.6888889}$$

$$= 1.3 \text{ (d.p)}$$

Hopefully this approach is less cumbersome. However the choice is yours!

Therefore $s = 1.3$ (1 d.p) and $s^2 = 1.69$ (2 d.p) as before!

Example 2

A class of 50 pupils obtained the following marks in a mathematics test.

Mark	1	2	3	4	5	6	7	8
Frequency	2	2	10	12	12	7	4	1

Calculate the standard deviation and variance of the marks.

Solution:

We shall use the formula

$$s = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

Tabulating the given data and required values for the formula we have:

Mark (x)	f	fx	x^2	fx^2
1	2	2	1	2
2	2	4	4	8
3	10	30	9	90
4	12	48	16	192
5	12	60	25	300
6	7	42	36	252
7	4	28	49	196
8	1	64	64	64

$\sum f = 50$ $\sum fx = 222$ $\sum fx^2 = 1104$

$$\text{Then } s = \sqrt{\frac{1104}{50} - \left(\frac{222}{50}\right)}^2$$

$$= \sqrt{\frac{1104}{50} - \frac{49284}{2500}}$$

$$= \sqrt{\frac{55200 - 49284}{2500}}$$

$$= \sqrt{\frac{5916}{2500}}$$

$$= \sqrt{3.3664}$$

$$= 1.54(2d.p)$$

$\therefore$ $s = 1.54$ (2 d.p)

and $s^2 = 2.37$ (2 d.p)

Use the formula = $\sqrt{\frac{\sum f\left(x-(\bar{x})\right)^{2}}{\sum f}}$ to find the standard deviation and

compare with the answer obtained above. Which formula is more friendly?

***A*CTIVITY: MT/2/2-16**

1. Find the standard deviation and variance for the number of heads obtained when coins were tossed 100 times, using the table of results below:

Number of heads	**Frequency**
0	4
1	15
2	34
3	29
4	16
5	2

2. The following marks were obtained by 50 students during an Agriculture Practical test.

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

Obtain a frequency distribution by means of a tally chart and calculate the standard deviation and variance of the scores.

3. In a swimming match, the times taken to the nearest second by 20 children to swim a given length were as follows:

31	27	24	26	31	25	26	32	27	31
26	32	30	32	29	25	29	27	26	32

Obtain a frequency distribution and find the variance and standard deviation of these times.

4. A population survey was made of the ages of the members of a Local Council 1 with the following results:

49	93	35	56	16	50	63	30	86	65
93	84	25	77	28	54	50	12	85	55
27	79	68	26	66	80	91	62	67	52
71	63	51	40	46	61	62	67	57	53
66	52	49	54	55	52	56	59	38	52
42	53	40	51	58	52	27	56	42	86
57	34	50	45	43	50	39	50	75	51
55	31	44	57	96	40	51	61	75	55
36	46	47	33	25	48	83	31	54	36
45	78	20	54	56	74	15	45	67	56

Form a frequency distribution and use it to find the standard deviation and variance.

✓ Check your answers with those given at the end of this unit.

(ii) **Grouped Frequency Distributions**

When working with grouped frequency distributions we shall use the same formulae as for ungrouped frequency distributions. However, it is important to remember that the mid-points for each group serve as the corresponding x values.

Example 1:

The table below shows the ages of the LC 1 Executive Committee members in a given parish.

Age	**Frequency**
30 – 34	4
35 – 39	9
40 – 44	14
45 – 49	8
50 – 54	5

Find the variance and standard deviation.

Solution:

We shall use s = $\sqrt{\dfrac{\sum fx^2}{\sum f} - \left(\dfrac{\sum fx}{\sum f}\right)^2}$

The table is as below:

Group	**x md pt**	**f**	**fx**	$\mathbf{x^2}$	$\mathbf{fx^2}$
30 – 34	32	4	128	1024	4096
35 – 39	37	9	333	1369	12321
40 – 44	42	14	588	1764	24696
45 – 49	47	8	376	2209	17672
50 – 54	52	5	260	2704	13520
		$\Sigma f = 40$	$\Sigma fx = 1685$		$\Sigma fx^2 = 72305$

Then s = $\sqrt{\dfrac{\sum fx^2}{\sum f} - \left(\dfrac{\sum fx}{\sum f}\right)^2}$

$$= \sqrt{\frac{72305}{40} - \left(\frac{1685}{40}\right)^2}$$

$$= \sqrt{\frac{72305}{40} - \frac{2839225}{1600}}$$

$$= \sqrt{\frac{52305}{1600}}$$

$$= \sqrt{33.109375}$$

$\therefore \quad = \quad$ 5.75 (2 d.p and $s^2 = 33.11$(2 d.p)

Now find s using the formula:

$$s = \sqrt{\frac{f\left(x-(\overline{x})\right)^2}{\sum f}}$$

Example 2:

The time in minutes taken by different people to complete a written interview is given in the table below:

Time t	38 - 42	43 - 47	48 - 52	53 - 57	58 - 62	63 - 67	68 - 72	73 - 77
Frequency	1	2	9	11	8	5	3	1

Calculate the standard deviation:

Solution:

Time t	**x md pt**	**f**	**fx**	**x^2**	**fx^2**
38 – 42	40	1	40	1600	1600
43 – 47	45	2	90	2025	4050
48 – 52	50	9	450	2500	22500
53 – 57	55	11	605	3025	33275
58 – 62	60	8	480	3600	28800
63 – 72	70	3	210	4900	14700
73 – 71	75	1			
		$\Sigma f = 40$	$\Sigma fx = 2275$		$\Sigma fx^2 = 131675$

Then using s $= \sqrt{\frac{fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$

$$= \sqrt{\frac{131675}{40} - \left(\frac{2275}{40}\right)^2}$$

$$= \sqrt{\frac{91375}{1600}}$$

$$= \sqrt{57.109375}$$

= 7.56 (2 d.p)

= 7.56 (2 d.p)

Once again, evaluate s for this example using the formula s $= \sqrt{\frac{\sum f\left(x = (\bar{x})\right)^2}{\sum f}}$

This should help you appreciate the advantage of both formulas.

ACTIVITY: MT/2/2-17

For each of the frequency distributions below, find the variance (s2) and the standard deviation (s).

1.

Number of mistakes made by students while translating from English to Swahili.	Number of Students
17 – 19	5
20 – 22	63
23 – 25	39
26 – 28	24
29 – 31	17
32 - 34	2

2. Lifetime of electric bulb

Hours	Frequency
400 – 450	22
451 – 501	38
502 – 552	62

3.

Age group	Number of Students
16 – 20	20
21 – 25	18
26 – 30	14
31 – 35	18
36 – 40	16
41 – 45	24

4.

Weight (kg)	Frequency
90 – 99	4
100 – 109	23
110 – 119	49
120 – 129	38
130 – 139	17
140 – 149	6
150 – 159	3

✓ Check your answers with those given at the end of this unit.

(c) USE OF CODING METHOD TO FIND THE STANDARD DEVIATION

From the exercises of Activity MT/2/2-5, you may have noticed how heavy the computations for s become as the values of x become larger and also the frequencies take on big values. Computation of the standard deviation of a frequency distribution is easiest calculated by using a coded method and an assumed mean. When this method is used, the standard deviation is found by using the formulae.

$$s = c\sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$$

Where d = the deviation from the assumed mean
f = the class frequency
c = class interval (or class width/size)

We shall use this formula for the examples earlier looked at to show that the methods agree.

Example 1:

The grade point averages (GPA) of 15 college graduates in section (b) examples were given as:

GPA:	1	2	3	4	5	6
f:	1	2	3	5	3	1

Use the coding method to calculate s.d.

Solution:

x	d	f	fd	d^2	fd^2
1	-3	1	-3	9	9
2	-2	2-	-4	4	8
3	-1	3	-3	1	3
4	0	5	0	0	0
5	1	3	3	1	3
6	2	1	2	4	4
		$\sum f = 15$	$\sum fd = {}^-5$		$\sum fd^2 = 27$

Assumed mean = 4
Class interval = 1

Then s $= \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$

$= 1\sqrt{\frac{27}{15} - \left(\frac{^-5}{15}\right)^2}$

$= \sqrt{\frac{27}{15} - \frac{25}{225}}$

$= \sqrt{\frac{380}{225}}$

$= \sqrt{1.6888889}$

s = 1.3 (1 dp) as before

Example 2:

The marks of 50 pupils in a Mathematics test were given in section (b) example 2 as:

Mark	1	2	3	4	5	6	7	8
Frequency	2	2	10	12	12	7	4	1

Use the coding method to find the standard deviation.

Solution:

x	d	f	fd	fd2
1	-3	2	-6	18
2	-2	2	-4	8
3	-1	10	-10	8
4	0	12	0	0
5	1	12	0	0
6	2	7	14	28
7	3	4	12	36
8	4	1	4	16
		$\Sigma f = 50$	$\Sigma fd = 22$	$\Sigma fd^2 = 128$

Assumed mean 4; c = 1

$$\text{Now have } s = c.\sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$$

$$= \quad 1.\sqrt{\frac{128}{50} - \left(\frac{22}{50}\right)^2}$$

$$= \quad \sqrt{\frac{5916}{2500}}$$

$$= \quad \sqrt{2.3664}$$

$\therefore$ s = 1.54 (2d.p) as before

Example 3:

Use the coding method to find s for the frequency distribution below:

Age	**Frequency**
30 – 34	4
35 – 39	9
40 – 44	14
45 – 49	8
50 – 54	5

Compare your answer with the one obtained in section b (ii) example 1.

Solution:

Age	**x mdpt**		**d**	**d^2**	**f**	**fd**	**df2**
30 – 34	32	-2	4		4	-8	16
35 – 39	37	-1	1		9	-9	9
40 – 44	42	0	0		14	0	0
45 – 49	47	1	1		8	8	8
50 – 54	52	2	4		$\underline{5}$	$\underline{10}$	$\underline{20}$
					$\Sigma f = 40$	$\Sigma fd = 1$	$\Sigma fd^2 = 53$

Assumed mean = 42; c = 5

Using the coded formula s = c

$$S = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$$

$$s = 5\sqrt{\frac{53}{40} - \left(\frac{1}{40}\right)^2}$$

$$s = 5\sqrt{\frac{2119}{1600}}$$

$$= 5\sqrt{1.324375}$$

$\therefore$ s = 5.75 (d.p) as found earlier on.

Example 4:

Find the standard deviation for the frequency distribution in section b (ii) Example 2 using the method of coding.

Solution:

Time t	x mdpt	d	f	fd	fd2
38 – 42	40	-3	1	-3	9
43 – 47	45	-2	2	-4	8
48 – 52	50	-1	9	-9	9
53 – 57	55	0	11	0	0
58 – 62	60	1	8	8	8
63 – 67	65	2	5	10	20
68 – 72	70	3	3	9	27
73 – 77	75	4	1	4	16
			$\Sigma f = 40$	$\Sigma fd = 15$	$\Sigma fd^2 = 97$

Assumed mean = 55;
and c = 5

Then $s = c.\sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$

$$= \quad 5.\sqrt{\frac{97}{40}-\left(\frac{15}{40}\right)^2}$$

$$= \quad 5.\sqrt{\frac{3655}{1600}}$$

$$= \quad 5 \times 1.5114149$$

$$\therefore s = \quad 7.56 \text{ (2 d.p)}$$

Once again the answer agrees with the one obtained earlier. Consequently, we have a choice to make from three different formulae.

ACTIVITY: MT/2/2-18

1. Work through the four examples of section (c) using a different assumed mean from the one in the example in each case. Compare your answers with the ones given.

2. Work through Activity MT/2/2-5, using the coding method to find the standard deviation. Compare your answers with those given at the end of this unit for Activity MT/2/2-5.

✓ Check your answers with those given at the end of this unit.

TOPIC 5: QUARTILES, PERCENTILES AND RANGE

In this topic, you will learn about:

- Methods of calculating quartiles and percentiles.
- Approximating quartiles and percentiles from cumulative frequency curve (ogive).
- Approximation of range and inter-quartile range from cumulative frequency curve.

(a) METHODS OF CALCULATING QUARTILES AND PERCENTILES

We have seen that the median divides a set of values into two equal parts. Using similar method, we can divide the set into any number of subsets each containing the same number of observations. ‘Quartiles’ or ‘fractiles’ is the general name given the values that divide up ranked (ordered) data into such subsets. In other words, quartiles are values below or above which given parts of data must fall.

(i) **Quartiles**

The values that divide a set into four equal parts are called quartiles. There are three quartiles that are often denoted by the symbols Q_1, Q_2 and Q_3.

Q_1 Is the first or lower quartile and is a value such that $\frac{1}{4}$ (or 25%) of the observations are less than or equal to it in value.

Q_2 Is the second quartile and the value of Q2 is equal to the median. $\frac{2}{4}$ of $\frac{1}{2}$ (or 50%) of the values are less than or equal to Q2.

Q_3 Is the third or upper quartile and is a value such that $\frac{3}{4}$ (or 75^) of the values are less than or equal to it.

(ii) **Percentiles**

The percentiles are the values which divide the set of observations into one hundred equal parts and these are denoted by P_1, P_2, P_3…, P_{99}.

Percentiles are such that:

1% of the values are less than or equal to	P_1
2%	P_2
3%	P_3
3%	P_3
3%	P_3
3%	P_3
3% and 99%	P_{99}

Example 1

For the information given in the table below, find the lower and upper quartiles. Find also the 60th and 90th percentiles.

Marks	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59
No. of Students	2	7	31	8	2

The first step is to obtain a cumulative frequency table.

Mark	**f**	**cf**
10 – 19	2	2
20 – 29	7	9
30 – 39	31	40
40 – 49	8	48
50 – 59	2	50

Next, we shall use the formula.

$$Q_1 = L + \frac{j}{f} \times c$$

To get the lower quartile where:

L	=	the lower boundary of the class into which the lower quartile must fall.
f	=	the frequency of the class into which the lower quartile must fall.
j	=	the number of items we still lack when we reach L and
c	=	the class interval.

Note that this formula is similar to the one used to obtain the median for grouped data.

Indeed you can write it as:

$$Q_1 = L + \frac{\left(\frac{N}{4} - Cf_b\right)}{f} c$$

Where $\frac{N}{4} - Cf_b = j$

and N = Total number of observations

cf_b = the cumulative frequency for the class below that into which the lower quartile must fall.

You will recall that Q_1 corresponds to $\frac{1}{4}$ of the total frequency.

In this example, therefore, Q1 corresponds to $\frac{1}{4} \times 50$ of the total frequency.

Clearly $\frac{1}{4} \times 50 = 12.5$

Examining the cumulative frequency column we see that Q1 must fall in the 30 – 39 class which has L = 29.5, f = 31. and cf_b = 9.

Substituting in the formula:

$$Q_1 = L + \frac{\left(\frac{N}{4} - Cf_b\right)}{f} c$$

We have $Ø_1 = 29.5 + \frac{(12.5 - 9)}{31} \times 10$

Do you realize that the class interval (or width) is 10? Right! Let us continue with our computation.

$$Q1 = 29.5 + \frac{3.5}{31} \times 10$$

$$= 29.5 + 1.120323$$

$$= 30.6 \text{ (1 d.p)}$$

Therefore the lower quartile $Q_1 = 30.6$

Similarly, the upper quartile Q3 can be got from the formula:

$$Q_3 = L + \left(\frac{\frac{3N}{4} - Cf_b}{f} \right) c$$

Where

L	=	the lower boundary of the class into which the upper quartile must fall.
f	=	the frequency of the class into which the lower quartile must fall.
N	=	the number of observations.
Cf_b	=	the cumulative frequency for the class below that into which the upper quartile must fall and
c	=	the class interval.

Again looking at our cumulative frequency table, $\frac{3N}{4} = \frac{3}{4} \times 50 = 37.5$, which

implies that $_3$ must fall in 30 – 39 class, which has L = 29.5, f = 31 and $cf_b = 9$. Again c = 10.

Substituting in the formula

$$Q_3 = L + \frac{\left(\frac{3N}{4} - Cf_b \right) c}{f}$$

$$\text{We have } Q_3 = 29.5 + \frac{(37.5 - 9)}{31} \times 10$$

$$= 29.5 + \underline{28.5 \times 10}$$

$$= 29.5 + \frac{28.5}{31}$$

$$= 29.5 + 9.1935484$$

$$= 38.7 \text{ (1 d.p)}$$

The upper quartile Q_3 $= 38.7$

The 60th percentile takes position $\frac{60N}{100}$

But $\frac{60N}{100} = \frac{60}{100} \times 50$

$$= 30$$

Using the same formulas as for the quartiles except for replacing $\frac{N}{4}$ or $\frac{3N}{4}$ by $\frac{60N}{100}$

we shall have: $P_{60} = L + \frac{\left(\frac{60N}{4} - Cf_b\right)c}{f}$

In which class does P_{60} lie?

Correct! It is in the 30 – 39 class just a Q_1 and Q_3 were

Hence, P60 $= 29.5 + \frac{(30 - 9)}{31} \times 10$

$$= 29.5 + \frac{210}{31}$$

$$= 36.3 \text{ (1 d.p)}$$

The 60th percentile $= 36.3$

Finally, P_{90} = $L + \dfrac{\left(\frac{90N}{100} - Cf_b\right)c}{f}$

We see that $\frac{90N}{100}$ = $\frac{90}{100} \times 50$

So where does our P_{90} lie?

Yes! It lies in the 40 – 49 class, which has L = 39.5, cf_b = 40, f = 8 and c = 5

We therefore have:

$$P_{90} = L + \frac{\left(\frac{90N}{100} - Cf_b\right)c}{f} \text{ becoming}$$

$$P_{90} = 39.5 + \frac{(45 - 40)}{8} \times 10$$

$$= 39.5 + 6.25$$

$$= 45.75$$

$$= 45.75$$

$$\therefore \quad P_{90} = 45.75$$

You should have noticed that the formulae for calculating quartiles and percentiles are similar to that of calculating the median. At this stage, read and work through all the examples and exercises on median in modules MT/1 and MT/2 topic 3. Go through the example above at least one more time. You can now proceed a lot more comfortably!

Example 2

When some young boys were asked the number of meals they had prepared during their December holidays, they gave the answers in the table below:

No. of meals	0 - 1	2 - 3	4 - 5	6 - 7	8 - 9
F	18	25	11	5	1

Find the upper quartile and the 78th percentile for this distribution.

Solution

No. of meals	f	cf
0 - 1	18	18
2 - 3	25	43
4 - 5	11	54
6 - 7	5	59
8 - 9	1	60

$$Q_3 =$$

$$\text{But } \frac{3N}{4} = 3 \times \frac{60}{4}$$

$$Q_3 = 45$$

$\therefore$ Q_3 lies in 4 - 5 class which has L = 3.5, Cfb = 43, f = 11 and c = 2.

$$\text{Then } Q_3 = L + \frac{(3N_{-c})}{}$$

$$= 3.5 + \frac{(45-43)}{11} \times 2$$

$$= 3.5 +$$

$$= 3.86 \text{ (2 d.p)}$$

$\therefore$ the upper quartile Q_3 = 3.86

$$P_{78} = L +$$

$$\text{But } \frac{78N}{100} = \frac{78 \times 60}{100}$$

$$P_{78} = 46.8$$

$\therefore$ P_{78} lies in the 4 – 5 class which has L = 3.5, cfb = 43, f = 11 and C = 2

Hence P78 = L +

$$= 3.5 + \frac{(46.8 - 43)}{11} \times 2$$

$$= 4.19 \text{ (2 d.p)}$$

$$\therefore \quad P_{78} = 4.19$$

ACTIVITY: MT/2/2-19

1. The frequency distribution below gives the heights of workers in a certain factory.

Height (cm)	156 – 158	159 – 161	162 – 164	165 – 167	168 – 170
Frequency	2	10	18	7	3

Find Q_1 Q_3 P_{15} P_{70} and P_{95}

2. The table below shows the annual incomes in US dollars for a group of 34 people.

Income ($)	1000 – 15000	15000 – 2000	2000 – 2500	2500 - 3000
No. of people	2	12	12	8

Calculate Q_1 Q_3 P_{50} and P_{80} to the nearest $

3. The marks obtained by a group of 100 pupils in an English Language Examination were as follows:

Mark	10 – 19	20 – 29	30- 39	40 – 49	50 – 59
No. of pupils	4	14	62	16	4

Calculate Q_2, Q_3, P_{25} and P_{66}

4. Calculate Q_2, P_{50} Q_3 and P_{75} for the distribution below:

Mass (kg)	0 – 4.9	5 – 9.9	10 – 14.9	15 – 15.9	20 – 24.9
Frequency	50	64	43	26	17

Comment about the answers you obtain.

✓ Compare your answers with the ones given at the end of this unit.

(b) APPROXIMATING QUARTILES AND PERCENTILE FROM THE CUMULATIVE FREQUENCY CURVE

Just like we approximated the median from the ogive, we can find approximate values of quartiles and percentiles from well constructed ogives. We are going to use the example in section (a).

Example

Find Q1, Q3, P and P using the ogive for the distribution below:

Marks	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59
No. of students	2	7	31	8	2

Solution

We first form a cumulative frequency table and then draw the ogive.

Mark	f	cf
10 – 19	2	2
20 – 29	7	9
30 – 39	31	40
40 – 49	8	48
50 – 59	2	50

We have chosen the following scales:

Vertical: 2 cm represents 10 units of cumulative frequency.
Horizontal 2 cm represents 10 marks.

A Cumulative Frequency Curve for the Marks obtained by 50 students

*

(c) From the calculation in section (a) we found that Q_1, is the 12.5^{th} in value. To find Q_1 from the ogive, we shall find 12.5 on the vertical scale and locate its corresponding value on the horizontal scale. This will be the value of Q_1.

Similarly Q_3 on the horizontal scale corresponds to a cumulative frequency value of 37.5. P_{60} corresponds to a c.f. of 30 while P_{60} corresponds to a c.f. of 45. From the ogive, we therefore have:

Q_1	=	30.5
Q_3	=	37.6
P_{60}	=	36.6
P_{90}	=	45.8

ACTIVITY: 2/2/-20

1. Construct an ogive for Example 2 Section (a) for the meals prepared by boys during their December holiday. From the ogive estimate, $\varnothing_3$ and P_{78}.

2. Work through Activity MT/2/2-6, using cumulative frequency curves in each case to find the required quartiles. Compare your answers with those given for Activity MT/2/2-7.

✓ Check your answers with those given at the end of this unit.

(d) APPROXIMATION OF RANGE AND INTERQUARTILE RANGE FROM CUMULATIVE FREQUENCY CURVE

The range is the difference between the largest and smallest observations in a distribution. That is, range = largest observation – smallest observations. The range gives some idea of the spread of a distribution but it depends solely upon the extreme values of the data. It is therefore seldom used as a measure for a frequency distribution.

The interquartile range is given by:

Interquartile range = $Q_3 - Q_1$

Where Q_3 is the upper quartile and Q_1 is the lower quartile.

The interquartile range ignores extreme values of a distribution. It is centred upon the median.

From the definitions, we see that these two measures are easy to get once the largest, smallest Q_1 and Q_3 of the observations are known.

Example

Using the ogive in section (b), $Q_1 = 30.5$
$Q_3 = 37.6$

Therefore interquartile range	=	$Q_3 - Q_1$
	=	$37.6 - 30.5$
	=	7.1
and range	=	largest observation – smallest observation
	=	50

***A*CTIVITY: MT/2/2-21**

Using the ogives constructed in Activity MT/3/2-7, estimate Q1, QØ3 in each case and find the interquartile range.

✓ Compare your answers with those given at the end of this unit.

TOPIC 6: APPLICATION OF STANDARD DEVIATION

When working with standard deviation, we observe that the dispersion of a set of data is small if the values are bunched closely about their mean and that it is large if the values are scattered widely about their mean. Consequently, we can now say that if the standard deviation of a set of data is small, the values are concentrated near the mean, and if the standard deviation is large, the values are scattered widely about the mean. The standard deviation can be used to compare measures from different distributions. We shall look at three different cases.

(a) The Coefficient of Variation

This is a measure of relative variation. It expresses the standard deviation as a percentage of what is being measured at least on the average. The coefficient of variation.

$$V = \frac{s}{x}100$$

Example

During recent months, the price of beef has averaged U.Shs 2400= per kilogram with a standard deviation of U.S Shs. 0.30= and the price of fish averaged U.Shs. 2800= per kilogram with a standard deviation of U.Shs.0.42=. For which of beef and fish is the price relatively more variable?

Solution

For beef, the coefficient of variation V $= \frac{s}{x}.100$

$$= \frac{0.30}{2400} x100$$

$$= 0.0125\%$$

For fish, V $= \frac{s}{x}.100$

$$= \frac{0.42}{2800} x100$$

$$= 0.015\%$$

The price of fish is therefore relatively more variable than that of beef.

(b) **Standard Scores**

Measures from different distributions can be compared by using standardized scores. A score in a distribution can be expressed in terms of the mean and standard deviation by using the formula.

$$Z = \frac{x - \bar{x}}{s}$$

In these standardized scores (or standard units) Z tells us how many standard deviations a value lies above or below the mean of the set of data to which it belongs.

Example 1

A student obtained 84 marks in a mathematics examination for which the mean mark was 70 with a standard deviation of 10. Calculate the standardized set score for this student.

Solution: x = 84, $x = 70$, s = 10

Therefore

$$Z = \frac{x - x}{S}$$

$$= \frac{84 - 70}{10}$$

$$= 1.4$$

Example 2

A girl obtains 70 marks in a final examination in Biology for which the mean mark was 65 with a standard deviation of 8 marks. In History, she obtained 68 marks but in this subject, the mean mark was 60 with a standard deviation of 10 marks. In which examination did she do better?

Solution:

In Biology, x = 70, $\bar{x}$ = 65 and s = 8

Therefore

$$z = \frac{70 - 65}{8}$$

$$= 0.625$$

In History, x = 68, $\bar{x}$ = 60 and s = 10

Therefore $z = \dfrac{68 - 60}{10}$

$= 0.8$

It can be seen that the girl had a higher standardized score in History and therefore she has done better in History than in Biology.

(c) **Skewness**

In your Professional Studies Module PS/1 you learnt that for a symmetrical distribution (i.e. a normal distribution), the arithmetic mean, the median and the mode all have the same value. However, if the distribution is skewed, the mean, median and mode all have different values (See Figures MT/ 2/1-1 and MT/2/2-2)

Skewness is the degree of a symmetry or a departure from symmetry for a distribution.

Symmetrical bell-shaped distribution

*

Figure Mt/2/2.1: For a systematical distribution, the arithmetic mean, median and mode have the same value.

Negative skewed distribution Positively skewed distribution

Figure MT/2/2.2: When a frequency distribution is skewed the mean, median and mode have different values.

For frequency distribution which are unimodal and moderately skewed, an appropriate relationship between the arithmetic mean, the median and the mode is:

$$\text{Mean} - \text{mode} = 3(\text{mean} - \text{median})$$

The degree of skewness is then given by Pearson's coefficient of skewness, SK which is calculated from the formula.

$$SK = \frac{3(\text{mean} - \text{median})}{\text{Standard deviation}}$$

This formula automatically gives the direction of the skew. If the result is positive, the skew is positive, if the result is negative, the skew is negative. It will be remembered that for positive skew, the longer tail is to the right and for negative skew, the longer tail is to the left. Note also from the Figure above the positions of mean and median for each case of skewness.

For positive skewness, the mean is greater than the median so that (mean – median0 is positive. Consequently, $SK = \frac{3(\text{men} - \text{median})}{\text{Standard deviation}}$ is positive.

In the case of negative skewness, the mean is less than the median, such that (mean – median) is negative, giving:

$$SK = \frac{3(\text{mean} - \text{median}) \text{ as negative}}{\text{Standard deviation}}$$

We also note that the higher the value of Pearson's coefficient of skewness, the greater the degree of skewness. For a perfectly symmetrical distribution, SK is zero and in general, it can take any value between $^{-}3$ and $^{+}3$.

Example 1

For a skewed distribution, the mean is 22, the median is 20 and the standard deviation is 4. Calculate Pearson's coefficient of skewness and sketch the distribution curve.

Solution

$$SK = \frac{3(\text{mean} - \text{median})}{\text{Standard deviation}}$$

$$= \frac{3(22 - 20)}{4}$$

$$= 1.5$$

This means that the distribution is positively skewed.

Example 2

A distribution has a mean of 35 and a median of 40, the standard deviation being 5. Calculate Pearson's coefficient of skewness and sketch the distribution curve.

Solution

$$SK = \frac{3(\text{mean} - \text{median})}{\text{Standard deviation}}$$

$$= \frac{3(35 - 40)}{5}$$

$$= .3$$

This means that the distribution is negatively skewed.

***ACTIVITY*: MT/2/2-22**

1. A student's marks for English and French were 48 and 60 respectively. In English the mean mark was 50 with a standard deviation of 20 and in French the mean mark was 62 with a standard deviation of 15. In which subject did the student do better?

2. On a final examination in statistics, the mean mark of a group of 50 students was 78 and the standard deviation was 8.0. In algebra, however, the mean final mark of the group was 73 and the standard deviation was 7.6. In which subject is the relative variation greater?

3. Find Pearson's coefficient of skewness for the following distribution and sketch the distribution:

(a)

Wages ($)	50 - 59	60 - 69	70 - 79	80 - 89	90 - 99	100 - 109	110 - 119
No. of Employees	8	10	16	14	10	5	2

(b)

Mark:	61	64	67	70	73
Frequency:	5	18	42	27	8

✓ Check your answers with those given at the end of this unit.

NOTES AND ANSWERS TO ACTIVITIES

ACTIVITY: MT/2/2-1

1. The highest mark is 94.
2. The lowest mark is 42.
3. The difference between the highest mark and the lowest mark is 52.

4.

40 – 44	45 – 49	50 – 54	55 – 59	60 – 64
65 – 69	70 – 74	75 – 79	80 – 84	85 – 89

5. There are no numbers in those classes.

ACTIVITY: MT/2/2-2

1. (i) The highest mark is 85.
   (ii) The lowest mark is 25.
   (iii) The range of the distribution is 60.

2. (i)

25 – 29	30 – 34	35 – 39	40 – 44	45 – 49
50 – 54	55 – 59	60 – 64	65 – 69	70 – 74
75 – 79	80 – 84	85 – 89 are the classes.		

   (ii) The last class is 85 - 89

ACTIVITY: MT/2/2-3

1.(a) Frequency table

Classes	**Tally**		**Frequency**
75 – 79	###		5
80 – 84	###	/	6
85 – 89	###	////	9
90 – 94	###	///	8
95 – 99	###	//	7
100 – 104	///		2

There are six classes formed.

(b) Modal class is 85 – 89.

2.(i)

Classes	Tally		Frequency 9f)
295 – 304	###	/	6
305 – 314	///		3
315 – 324	###	//	7
325 – 334	//		2
335 – 344	###		5
345 – 354	###	###	10
355 – 364	//		2

(ii) The Modal class is 345 – 354

ACTIVITY: MT/2/2-4

The class limits for the classes in table 1 are as follows:

Class	**Class limits are:**	
40 – 49	40 lower class limit	49 upper class limit
50 – 59	50 " " "	59 " " "
60 – 69	60 " " "	69 " " "
70 – 70	70 " " "	79 " " "
80 – 89	80 " " "	89 " " "
90 – 99	90 " " "	99 " " "

ACTIVITY: MT/2/2-5

1.(i) Consider a class of the same interval just before 70 – 79 is 60 – 69, then lower class boundary for 70 – 79 is $\frac{69 + 70}{2} = 69.5$

Also consider a class of the same interval just above 70 – 79 which is 80 – 89, then upper class boundary is $\frac{79 + 80}{2} = 79.5$

(ii) Similarly the class boundaries for the class 90 – 99 are 89.5 and 99.5

2. (i) Average $= \frac{69.5 + 79.5}{2} = \frac{149}{2} = 74.5$

3. (i) Average $= \frac{70 + 79}{2} = \frac{149}{2} = 74.5$

   (ii) $\frac{90 + 99}{2} = \frac{189}{2} = 94.5$

4. The results in NOs. 2 and 3 are equal.

5. Class mark $= \frac{\text{Lower class boundary + upper class boundary}}{2}$

$$= \frac{\text{Lower class limit + upper class limit}}{2}$$

ACTIVITY: MT/2/2-6

1. (i) The highest score is 71
   (ii) The lowest score is 35

2. Frequency table

Classes	**Talley**			**Frequency 9f)**
35 – 39	///			3
40 – 44	//			2
45 – 49	~~////~~			5
50 – 54	~~////~~	~~////~~		10
55 – 59	~~////~~	~~////~~	//	12
60 – 64	~~////~~	~~////~~	/	11
65 – 69	~~////~~			5
70 – 74	//			2

To draw the histogram we need class boundaries.

Class		**Boundaries**	**Frequency (f)**
34.5	-	39.5	3
39.5	-	44.5	2
44.5	-	49.5	5
49.5	-	54.5	10
54.5	-	59.5	12
59.5	-	64.5	11
64.5	-	69.5	5
69.5	-	74.5	2

Histogram, see the graph.

*

ACTIVITY: MT/2/2-7

Classes	**Midpoints**	**Frequency 9f)**
31 – 35	33	2
36 – 40	38	6
41 – 45	43	5
46 – 50	48	7
51 – 55	53	25
56 – 60	58	49
61 – 65	63	33
66 – 70	68	8
71 - 75	73	3
76 – 80	78	4

ACTIVITY: MT/2/2-8

Frequency distribution table

1\.

Classes	Tally	Frequency 9f)
50 – 59	/	1
60 – 69	///	3
70 – 89	~~////~~	7
80 – 89	///	3
90 – 99	/	1

Cumulative frequency distribution table

2\.

Classes	less than	frequency	Cumulative frequency
50 – 59	59.5	1	1
60 – 69	69.5	3	4
70 – 79	79.5	7	11
80 – 89	89.5	3	14
90 – 99	99.5	1	15

Cumulative frequency curve (ogive). See the graph

Cumulative frequency curve (ogive)

*

ACTIVITY: MT/2/2-9

Age	**Class marks** (x)	**Frequency** (f)	**d = x - A**	**fd**
13 – 15	14	6	-3	.18
16 – 18	17	8	0	0
19 – 21	20	20	3	60
22 – 24	23	4	6	24
25 – 27	26	2	9	18
		$\sum f = 40$		$\sum fd = 84$

$$\text{Mean} = \text{assumed mean} + \frac{\sum fd}{\sum f}$$

$$= 17 + \frac{84}{40}$$

$$= 17 + \frac{21}{10}$$

$$= 17 + 2.1$$

$$= 19.1$$

ACTIVITY: MT/2/2-10

Class	**Class marks**	**u**	**f**	**fu**
110 – 114	112	-3	-5	-15
115 – 119	117	-2	-6	-12
120 – 124	122	-1	-17	-17
125 – 129	127	0	21	0
130 – 134	132	1	19	19
135 – 139	137	2	16	32
140 – 144	142	3	12	36
145 – 149	147	4	4	16

$$\text{Mean} = \text{The class mark with coding No. '0'} + \frac{\sum fu \times \text{class interval}}{\sum f}$$

$$= 127 + \frac{59 \times 5}{100}$$

$$= 129.95$$

ACTIVITY: MT/2/2-11

Age	Class boundaries	Frequency
15 – 19	14.5 – 19.5	22
20 – 24	19.5 – 24.5	35
25 – 29	24.5 – 29.5	13
30 – 34	29.5 – 34.5	9
35 – 39	34.5 – 39.5	24
40 – 44	39.5 – 44.5	30
45 – 49	44.5 – 49.5	40
50 – 54	49.5 – 59.5	19
55 – 59	54.5 – 59.5	15
60 – 64	59.5 – 64.5	10
65 – 69	64.5 – 69.5	9
70 – 74	69.5 – 74.5	8
75 – 79	74.5 – 79.5	5

(i) For the histogram, see the graph below.
Mode from the histogram = 44.5 + 1.5 = 46.0

(ii) Mode using the formula

$$= L_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} C$$

$$= 44.5 + \frac{(10)5}{10 + 21}$$

$$= 46.1$$

ACTIVITY: MT/2/2-12

Cumulative frequency distribution

1.

Age	less than	frequency	Cumulative frequency
15 – 19	19.5	22	22
20 – 24	24.5	35	57
25 – 29	29.5	13	70
30 – 34	34.5	9	79
35 – 39	39.5	24	103
40 – 44	44.5	30	133
45 – 49	49.5	40	173
50 – 54	54.5	19	192
55 – 59	59.5	15	207
60 – 64	64.5	10	217
65 – 69	69.5	9	226
70 – 74	74.5	8	234
75 – 79	79.5	5	239

*

2.(i) To find the median on a cumulative frequency curve N = 239
Cumulative frequency for the median is $\frac{239}{2} = 119.5$

Median = $39.5 + 2.55$ = 42.05

(ii) Median = $L + \left(\dfrac{\frac{N}{2} - B}{f} \right) C$

The definitions of these letters are in the notes.

Median =

$= 39.5 + \dfrac{119.5 - 103}{20}$

$= 39.5 + \dfrac{16.5}{20}$

$= 39.5 + 0.825$

$= 40.325$

ACTIVITY: MT/2/2-13

1. 2.3
2. 8.5
3. 10.2
4. 9.214

ACTIVITY: MT/2/2-14

See answers to MT/2/2-13

ACTIVITY: MT/2/2-15

1. 5.3
2. 72.3
3. 104.0
4. 84.7

ACTIVITY: MT/2/2-16

1. $s = 1.1$ $s^2 = 1.2$

2.

x	Tally	f
1	//	2
2	//	2
3	/// /// //	12
4	/// ///	10
5	/// /// //	12
6	/// //	7
7	////	4
8	/	1

$s = 1.6$ $s^2 = 2.56$

3.

x	Tally	f
24	/	1
25	//	2
26	////	4
27	///	3
29	//	2
30	/	1
31	///	3
32	////	4

4. $s = 2.7$ $s^2 = 7.3$

ACTIVITY: MT/2/2-17

1. $S = 3.4$ $S^2 = 11.5$
2. $S = 38.9$ $S^2 = 1511.3$
3. $S = 9.0$ $S^2 = 81.5$
4. $S = 12.3$ $S^2 = 150.7$

ACTIVITY: MT/2/2-18

See answers for MT/2/2-18

ACTIVITY: MT/2/2-19

1. $Q_1 = 0.9$ $Q_2 = 164.5$ $P_{15} = 159.7$
   $P_{40} = 2.2$ $P_{70} = 164.2$ $P_{95} = 168.5$

2. $Q_1 = 1770$ $Q_3 = 2479$ $P_{50} = 21245$
   $P_{80} = 2574.5$

3. $Q_2 = 34.7$ $P_{25} = 30.6$ $Q_3 = 38.7$

4. Q_2 = P50 = 8.4
Q_3 = P75 = 13.7

The median = the second quartile = 50^{th} percentile and the upper quartile = 75^{th} percentile.

ACTIVITY: MT/2/2-20

See the answers for MT/2/2-19

ACTIVITY: MT/2/2-21

- For meals, interquartile range = 2.7
- For heights of workers, interquartile range = 3.6
- For annual incomes, interquartile range = 709
- For marks in English Language, interquartile range = 8.1
- For mass (kg) interquartile range = 8.9

ACTIVITY: MT/2/2-22

1. English, standard score is higher.
2. Algebra
3. (a) SK = 0.13 (b) SK = 0.52

   Both are positively skewed.
   Sketches are more or less alike as below:

SUMMARY

In this unit, we have studied frequency distributions of grouped data and their graphical representations. The use of histograms, frequency polygons and cumulative frequency curves (ogives) were discussed.

We have seen how to use the diagrams for the frequency distributions to find measures of central tendency for example mode and median. We have also seen how to find these measures using formulae.

We have found out how to find the mean, another measure of central tendency using the assumed mean and the coding method.

You have broadened your knowledge of statistics through learning about the standard deviation and variance. You also looked at quartiles, percentiles and range as other measures of dispersion. Finally, you learnt about some applications of standard deviation.

What we have covered in this unit is useful in research and planning. We are hopeful you will be able to apply statistics for your research.

GLOSSARY

Raw data :	This is the information that we collect and not yet classified.
Class boundaries :	These are numbers that fall halfway between the top of one class and the bottom of the next.
Class size (Class interval):	This is the upper boundary minus the lower boundary for the class.
Class work:	This is the midpoint of the class interval.
Tallies:	These are strokes corresponding to the frequency for a particular class or quantity. They are usually in bundles of five.
Frequency:	The number of times a certain figure or quantity occurs in a given data.
Frequency distribution:	This is a table consisting of classes of the collected data with respective frequencies.
Mode of frequency distribution:	This is the quantity with the highest frequency.
Modal class:	This is the class where the ode falls.
Median class:	This is the class where the median falls.
Cumulative frequency:	This is the running total of consecutive frequencies.
Histogram:	This is a bar chart in which the areas of the rectangles are proportional to the frequencies.
Coding method:	Shorter and simpler method for calculating standard deviation, obtained by using one of the x values as an assumed mean.
Coefficient of variation:	A quicker method of calculating standard deviation using raw scores and avoiding the estimation of the mean.
Dispersion:	The spread of variation within a distribution.
Interquartile range:	Upper quartile – lower quartile.

Pearson"s coefficient of skewness: A measure of skewness of a distribution given by *3(mean – median) standard deviation.*

Percentiles: Values which divide a set of observations into equal parts. They are also known as fractiles.

Range: Difference between the largest and smallest values.

Skewness: The degree of departure from symmetry of a distribution.

Standard deviation: A measure of how observations are spread about their mean.

Standardized scores: Measures that tell us how many standard deviations a value lies above or below the mean of the data set to which it belongs.

Variance: The square of the standard deviation

END OF UNIT ASSIGNMENT

1. The following are the marks obtained by 45 students in a Mathematics examination at King's College Buddo.

72	76	82	76	59	67	73	80	90	63
78	73	91	66	88	63	74	86	67	71
51	86	82	76	48	72	87	72	86	76
49	96	71	72	85	87	71	83	66	71
51	46	81	82	64					

(i) Make a frequency table starting with classes 45 – 49, 50 – 54……..
(ii) Find the class interval.
(iii) Estimate the median mark.
(iv) Draw a histogram and use it to estimate the mode mark.

2. In a class of 40 pupils, the following marks were scored in a mathematics test (marked out of 25).

2	5	7	10	5	11	12	14	9	14
4	9	11	9	16	2	2	7	5	16
6	15	16	3	20	8	19	4	2	10
11	14	15	6	20	19	14	14	8	24

(a) Make a frequency distribution table with five classes, 1 – 5, 6 – 10,…….., 21 – 25.

(b) Construct a cumulative frequency distribution table and draw the ogive.

3.

Heights	20 - 22	23 - 25	26 - 28	29 - 31	32 - 34	35 - 37
Frequencies	1	12	16	13	6	2

The table above shows the length in cm of 50 seedlings.

(i) State the modal class.
(ii) Calculate the modal height.
(iii) Draw an ogive.
(iv) Use the ogive to estimate the median height.

4. In a class of 40, the teacher needed to estimate the mean age of the students.

Age	**Frequency**
13 – 15	6
16 – 18	8
19 – 21	20
22 – 24	4
25 – 27	2

(i) Using 17 as the assumed mean, calculate the mean age.
(ii) Find also the mean using the coding method.

5. At Kawanda Agricultural Research Institute, the heights of two-year old trees are measured to the nearest centimeter.

Height interval H(cm)	No. of trees f	Class mark x
120 – 124	6	
125 – 124	15	
130 – 134	40	
135 – 139	30	
140 – 144	9	

(a) Copy and complete the table. Work out the class mark for each clas.
(b) Using assumed mean 132, find the mean for the data.
(c) Why should the mean of grouped data be an estimate?

6. Given below are the masses of 8 young people measured to the nearest kilogram. Calculate the variance and standard deviation.

40, 49, 43, 35, 42, 43, 46, 36,

7. The hand spans of 40 children were measured to the nearest cm. RThe results are given below:

15	19	20	16	18	21	23	17	17	18
19	17	18	20	16	13	22	19	18	18
18	19	19	20	17	18	20	22	19	17
18	17	17	16	19	25	15	16	19	18

(a Form a frequency table (ungrouped) of the distribution.

(b) Find the mean, variance and standard deviation.

(c) Find Q_1, Q_2, Q_3, P_{10}, P_{60} and P_{80}

(d) What is the coefficient of variation for the distribution?

8. The table below shows the distribution of the scores obtained by 200 pupils in a Geography test.

Score	**No. of Pupils**
76 – 80	2
71 – 75	5
66 – 70	9
61 – 65	19
56 – 60	32
51 – 55	55
46 – 50	38
41 – 45	21
36 – 40	10
31 – 35	6
26 – 30	3

(a) Using the method of coding, calculate the standard deviation and variance.

(b) What is the range?

(c) Construct a cumulative frequency cure and use it to estimate.

(i) The median score.

(ii) The lower quartile, upper quartile and interquartile range

(iii) The number of students scoring more than 42.

(iv) The 60^{th} percentile.

(v) The percentile corresponding to a score of 68.

9.(a) Eleven people were asked to guess the mass of a cake to the nearest half kilogram. The results were:

2½, 1½, 2, 2½, 1½, 2½, 1, 1½, 2½, 1 kg

State (i) the median value

(ii) the range

(iii) the lower and upper quartiles

(b) Calculate the mean and standard deviation of the number 8, 11, 12, 14, 15 and 18. Correct each of these numbers into a standard score. Estimate the skewness of the distributions.

LEARNING OUTCOMES SELF-CHECKING EXERCISE

You have now completed Unit 2 of Module MT/2. The learning outcomes are listed below:You are expected to tick the column that reflects your learning.

	Learning Outcomes	***Satisfactory***	***Not Sure***
1.	I can write down groups/classes of the raw information/data accordingly.		
2.	I can tabulate the grouped data.		
3.	I can draw histograms and cumulative frequency curves (ogives) using the tabulated data.		
4.	I can draw frequency polygons using the tabulated data.		
5.	I can calculate the mean of the tabulated data using assumed mean and the method of coding.		
6.	I can compute the mode using the histogram.		
7.	I can compute the median using the cumulative frequency curve (ogives)		
8.	I can calculate the mode and median for grouped data using formulae.		
9.	I can name the different measures of dispersion.		
10.	I can calculate the standard deviation and variance from raw data.		
11.	I can calculate the standard deviation and variance from frequency distributions.		
12.	I can use the coding method to find the standard deviation.		
13.	I can calculate quartiles and percentiles.		
14.	I can approximate quartiles, percentiles and interquartile range from the ogive.		
15.	I can apply the knowledge of quartiles and percentiles to solving statistical problems.		

If you have placed a tick in the column for not sure, please go back to the information in the unit and reinforce your learning.

ANSWERS TO END OF UNIT ASSIGNMENTS: MT/2-3

1. (i)

Classes	Tally	Frequency (f)
45 – 49	///	3
50 – 54	//	2
55 – 59	/	1
60 – 64	///	3
65 – 69	////	4
70 – 74	~~////~~ ~~////~~ /	11
75 – 79	~~////~~	5
80 – 84	~~////~~ /	6
85 – 89	~~////~~ //	7
90 – 94	//	2
95 – 99	/	1

(ii) Class interval = 50 – 45 = 5

or = 49.5 - 44.5 = 5

(iii) Cumulative frequency distribution.

Classes	Frequency	Cumulative frequency
45 – 49	3	3
50 – 54	2	5
55 – 59	1	6
60 – 64	3	9
65 – 69	4	13
70 – 74	11	24
75 – 79	5	29
80 – 84	6	35

85 – 89	7	42
90 – 94	2	44
95 – 99	1	45

$$\text{Median} = L_1 + \frac{\left(\frac{N}{2} - B\right)}{f}$$

Where L = 69.5

N = 22.5

B = 13

F = 24 – 13

= 11

$$\text{Median} = 69.5 + \frac{22.5 - 13}{11}$$

$$= 69.5 + \frac{9.5}{11}$$

= 70.4

Class boundaries	Frequency
44.5 - 49.5	3
49.5 – 54.5	2
54.5 – 59.5	1
59.5 – 64.5	3
64.5 – 69.5	4
69.5 – 74.5	11
74.5 – 79.5	5
79.5 – 84.5	6
84.5 – 89.5	7
89.5 – 94.5	7

94.5 – 99.5		1

Mode = 69.5 + 2.5

= 72

Frequency distribution table

2.(a)

Classes	Tally	Frequency
1 – 5	~~////~~ ~~////~~	10
6 – 10	~~////~~ ~~////~~ /	11
11 – 20	~~////~~ ~~////~~ /	11
16 – 20	~~////~~ //	7
21 – 25	/	1

3. (i) Modal class is 26 – 28

(ii)

Heights	Class boundaries	Frequency
20 – 22	19.5 – 22.5	1
23 – 25	22.5 – 25.5	12
26 – 28	25.5 – 28.5	16
29 – 31	28.5 – 31.5	13
32 – 34	31.5 – 43.5	6
35 – 37	34.5 – 37.5	2

Mode = $L1 + \frac{(\Delta_1)c}{\Delta_1 + \Delta_2}$

L1 = $25.5 + \frac{(16-12)3}{16-13}$

= $25.5 + \frac{4}{3}x3$

= $25.5 + 4$

= 29.5

(iii) Cumulative frequency table

Heights	**Less than**	**Frequency**	**Cumulative frequency**
20 – 22	22.5	1	1
23 – 25	25.5	12	13
26 – 25	28.5	16	29
29 – 31	31.5	13	42
32 – 34	34.5	6	48
35 – 37	37.5	2	50

Cumulative frequency for the median is $\frac{50}{2} = 25$

Median is 25.5 + 7.5 x 0.3
Having used the scale
Every tiny square represents 0.3

Median = 25.5 + 2.25
= 27.75

4. (i)

Age	Class mark	x – A	f	f(x – A)
13 – 15	14	-3	6	-18
16 – 18	17	0	8	0
19 – 21	20	3	20	60
22 – 24	23	6	4	24
25 – 27	26	9	2	18
		$\Sigma f = 38$		$\Sigma f(x - A) = 84$

Assumed mean (A) = 17

Mean = Assumed mean + $\frac{\sum f(x - A)}{\sum f}$

= $17 + \frac{84}{38}$

= 19.2

(ii) By the method of coding

Age	**Class mark**	**Coding Numbers (u)**	**f**	**fu**
13 – 15	14	$^{-}2$	6	-18
16 – 18	17	$^{-}1$	8	0
19 – 21	20	0	20	60

22 – 24	23	1	4	24
25 – 27	26	2	2	18

$\Sigma f = 38$ $\Sigma fu = {}^{-}12$

$$\text{Mean} = 20 + \frac{12}{38} \times 3$$

$$= 19.05$$

5.

	Height interval H (cm)	No. of trees (f)	Class mark (x)
(a)	120 – 124	6	122
	125 – 129	15	127
	130 – 134	40	132
	135 – 139	30	137
	140 – 144	9	139

(b)

Class mark x - A (x)	x – A	(f)	f(x – A)
122	-10	6	-16
127	-5	40	-75
132	0	40	0
137	5	30	150
142	40	9	90

$\Sigma f = 100$ $\Sigma f(x - A) = 105$

Assumed mean = 132

$$\text{Mean} = 132 = \frac{105}{100}$$

$$= 132 + 1.05$$

$$= 133.05$$

(c) Mean of grouped data is just an estimate because it is the mean got from the class groups and not the actual heights of the trees.

6. Variance S^2 = 19.4 (1 d.p)
Standard deviation S = 4.4 (1 d.p)

7.(a)

x	Tally	f
13	/	1
15	//	2
16	////	4
17	~~////~~ //	7
18	~~////~~ ////	9
19	~~////~~ ///	8

20	////	4
21	/	1
22	//	2
23	/	1
25	/	1

(b) $x = 18.325$
$S^2 = 5.0$ (1 d.p)
$S = 2.2$ (1 d.p)

(c) $Q_1 = 17$
$Q_2 = 18$
$Q_3 = 19$
$P_{10} = 6$
$P_{60} = 19$
$P_{80} = 20$

(d) 27.3%

8. (a) $S = 9.4$ (1 d.p)
$S^2 = 87.4$ (1 d.p)

(b) Range = 55
(c) (i) 52.5

(ii) $Q_1 = 46.8$
$Q_3 = 58.2$
$Q_3 - Q_1 = 11.4$

(iii) 54.3

(iv)

9.(a) (i) 2
(ii) 1½
(iii) $Q_1 = 1½$
$Q_3 = 2½$

(b) $x = 13$
$S = 3.2$ (1 d.p)

X	Standard Score
8	-1.5625
11	-0.625
12	-0.3125
14	+0.3125
15	+0.625
18	+1.5625

SK = 0. The distribution is not skewed. It is normally distributed.

UNIT 3

PROBABILITY

INTRODUCTION

You are welcome to Unit 3 of Module 2, which is an extension of Probability. Look out for an enjoyable time as you learn more about probability, building on what you learnt in Module MT/1.

AIMS

The aims of this unit are:

1. To increase your knowledge of probability theory.
2. Enable you solve probability problems using a variety of approaches.
3. Enable you teach probability very well in the primary schools, using every day life examples.

SPECIFIC OBJECTIVES

By the time you have worked through this unit, you should be able to:

1. Solve probability problems using set theory.
2. Solve probability problems using probability trees.
3. Identify mutually exclusive and independent events.
4. Solve problems involving mutually exclusive and independent events.
5. Define a discrete probability distribution.
6. Find the expectation of a probability distribution.
7. Find the variance of a probability distribution.

UNIT ORGANISATION

This unit is organized into 3 topics.

Topic 1: Probability using set theory.

It has five sub-topics:

(a) The Universal set ε and sample space S.
(b) The subset A of a universal set ε and the event E in a sample space S.
(c) Representing probabilities in Venn diagrams.
(d) Solving probability problems using set theory.
(e) Sample space diagrams.

Topic 2: Mutually Exclusive and Independent Events

It has two sub-topics:

(a) Mutually exclusive events.
(b) Independent events

Topic 3: Probability trees and their use in solving problems

It has three sub-topics.

(a) Tree diagrams.
(b) Probability trees
(c) Using probability trees to solve problems.

SUBJECT ORIENTATION

Before you begin studying this unit, you should revise set theory and the simple probability of an event in Module MT/1. You should also be familiar with probability in the Primary School Syllabus.

STUDY REQUIREMENTS

❒ You will need a pen, pencil, ruler and an exercise book. You also need a quiet room free from internal and external disturbances.

Enjoy learning more about Probability.

TOPIC 1: PROBABILITY USING SET THEORY

In this topic, you are going to learn about:

- The Universal Set ε and the sample space, S.
- The sub-set A of a Universal set ε and the event e in a sample space S.
- Representing probabilities in Venn diagrams.
- Solving probability problems using set theory.
- Sample space diagrams.

(a) The Universal Set ε and the sample space, S

You have worked with the idea of the probability of an event and evaluated simple cases by listing and counting all the possible outcomes of a trial, then finding the proportion of outcomes in which the event occurs. This set of all possible outcomes of a trial is the Universal set, ε. Here, we shall call this set the sample space, S. Below are three example of sample space.

Example 1

When a coin is tossed once, it can land either heads (H) up or tails (T) up. The set of all possible outcomes is S = {H, T}.

Example 2

If you are to choose one of the colours of the rainbow, you will have as the sample space S = { Red, Orange, Yellow, Green, Blue, Indigo, Violet }.

Example 3

During the 1996 Uganda Presidential Elections, voters had S = {Museveni, Semogerere, Mayanja} as the sample space.

As you will notice, a sample space represents all possible outcomes of a specific trial. It will therefore vary from one trial to another. The choice of the sample space S depends on the information known about the details of the trial and on the questions we wish to answer.

ACTIVITY: MT/2/3-1

- Write down 5 different examples of sample spaces.
- Compare your answers with those at the end of this unit.

✓ Check your answers with those given at the end of this unit.

As you have already learnt from the previous sub-topic, the Universal set is what we refer to as the sample space when dealing with probability.

A set A, whose every element is a member of the Universal set ε is said to be a sub-set of ε. The sub-sets of our sample space S will be called events, E. We can use any other letter of the alphabet to refer to an event.

Consequently, we can refer to n(S), n(E) and so on, just as we do for sets. Let us now look at some examples of events.

Example 1

When a coin is tossed once, S = {H, T}. See topic (a) example 1. If we are interested in getting a head H, then our event, E = {H}.

Example 2

If we now toss two coins simultaneously, what is our sample space going to be? We can have:

both landing with heads up, HH;
both landing with tails up, TT;
the first landing with H up and the second landing with T up, HT, and the first landing with T up and the second landing with H up, TH.

This gives S = {HH, TT, HT, TH}

☞ **Note**: The coins are tossed simultaneously and we can distinguish them as first coin and second coin by a small coloured dot on one of them. If we are interested in at least one of the coins landing head, H up, then E = {HH, HT, TH} is our event.

Example 3

An integer is chosen from the numbers, 1, 2, 3,…, 10. if we are interested in even numbers, what will our set E be? What if we were interested in multiples of 3?

In this case, S = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
For even numbers, E = {2, 4, 6, 8, 10}
For multiples of 3, if we call the event A, then A = {3, 6, 9}.

Do you notice that the events, set varies according to the information we have about the event?

Write down the event set P, if P is choosing a prime number from the given sample space above.

ACTIVITY: MT/2/3-2

For the sample spaces you wrote down in Activity MT/2-1, write down at least three different events for each one of them.

✓ Check your answers with those given at the end of this unit.

(c) Representing Probabilities in Venn Diagrams.

Before we can represent probabilities in Venn diagrams, let us remind ourselves of two important points.

(i) When dealing with the probability of a simple event E, we defined the probability of E as $P(E) = \frac{n(E}{n(S)}$

$$\text{or } P(E) = \frac{\text{number of favourable outcomes}}{\text{number of total outcomes}}$$

Where E is the set of the event of our interest and S is the sample space or Universal set.

Now suppose our interest is getting the entire sample space S,

then $P(S) = \frac{n(S)}{n(S)} = 1$

In other words, the probability for the entire sample space is always 1. This is a very important point when we are representing probabilities in Venn diagrams. The sum of all probabilities in the diagram must be 1.

(i) For any event A, $0 \leq P(A) \leq 1$.

We can then represent probabilities in Venn diagrams, using the following points.

1. The sample space is equal to the Universal set. It will therefore be represented by a rectangle.

2. An event is a student of a sample space. It will be represented by a circle, within the rectangle or by part of a circle. It may also be for region outside the circles but within the rectangle.

3. We shall write the probability of an event in the region representing the event on the Venn diagram, just like we would write the number of elements for a set.
4. The total sum of probabilities in all the regions of the Venn diagram is 1, (representing the entire sample space).

Carefully study the following examples showing how to represent probabilities in Venn diagrams.

Example 1

When a coin is tossed once and we are interested in getting a head H, then:

$$S = \{H, T\}$$
$$E = \{H\}$$

What is the complement of E?

You got it right, $E' = \{T\}$

Do you realize that $n(S) = 2$, $n(E) = 1$ and $n(E') = 1$?

Then $P(E) = n(E) = \frac{1}{2}$ and $P(E) = \frac{1}{2}$

Similarly, since $S = E + E'$

Then $n(S) = n(E) + n(E')$

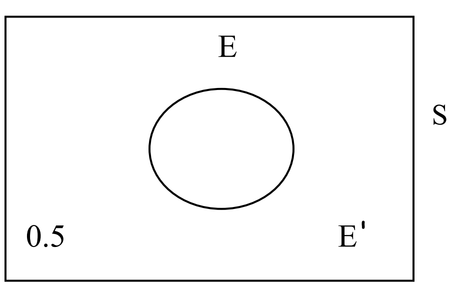
And $n(S) = n(E) + n(E')$
$n(S) = n(S) + n + n(S)$

Give us $1 = P(E) + P(E')$. This can be reorganized to give us:

$$\boxed{P(E') = 1 - P(E)}$$

The event E' is referred to as ‘not E'.

We can now draw a Venn diagram of probabilities as shown below. Remember the total sum of probabilities in all the regions of the Venn diagram is 1.



Example 2

An integer is chosen from the numbers, 1, 2, 3, …, 10. If A is the event that the integer chosen is an even number and B is the event that the chosen integer is a multiple of 3.

(i) Represent the information in a Venn diagram.

(ii) Find the corresponding probabilities for each event and write it in the corresponding region of the Venn diagram.

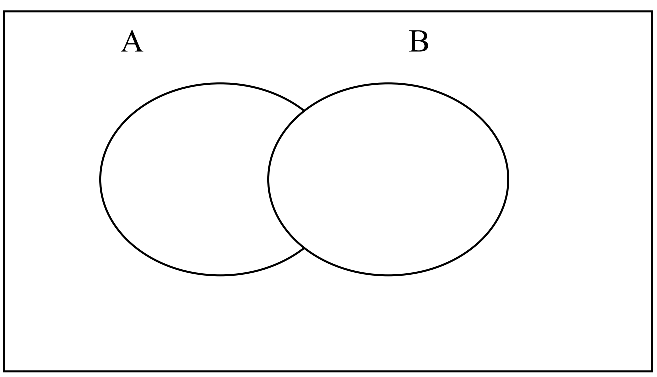
Solution

In this case,

$$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\},$$
$$A = \{2, 4, 6, 8, 10\} \text{ and}$$
$$B = \{3, 6, 9\}$$

What do you notice about sets A and B?
Do they have any elements in common? Yea!

Therefore, we have $A \cap B = \{6\}$. The following Venn diagram represents the given information.



S

(ii) $n(S) = 10$
$n(A) = 5$
$n(B) = 3$

Therefore:

$$P(A) = \frac{5}{10} = 0.5 \quad \text{and} \quad P(B) = \frac{3}{10} = 0.3$$

Note carefully!

(a) Since $n(A \cap B) = 1$

Then $P(A \cap B) = \frac{1}{10}$

$= 0.1$

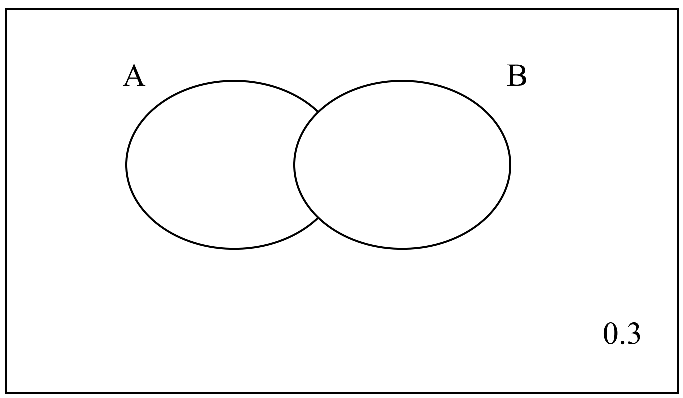
$A \cap B$ represents the event 'A and B', that is A and B occurring together.

(b) If we call the region outside the circles but within the rectangle D, then $n(D) = 3$.

Therefore $P(D) = \frac{3}{10} = 0.3$

We can then fill in the probabilities, just like we did for the integers.

Remember, we work from the centre outwards. Our Venn diagram will therefore look like this:



Do you realize that $0.4 + 0.1 + 0.2 + 0.3 = 1$?

Study the venn diagram above and write down the values of $P(A \cup B)$ and $(A' \cap B')$.

$A \cup B$ is read as A or B and refers to "event A or event B or both. What do you think $A' \cap B'$ means? Which region (s) of the venn diagram represents $A' \cap B'$?

Example 3

A die has its faces marked with the numbers 1, 2, 3, 4, 5 and 6. It is tossed once and the number on its top face recorded. If B is the event that the top number is a prime and D is the event that it is a 6, draw a Venn diagram showing the probabilities for each region for the given information.

Solution

We have $S = \{1, 2, 3, 4, 5, 6\}$ and $n(S) = 6$,

$B = \{2, 3, 5\}$ and $n(B) = 3$,

$D = (6)$ and $n(D) = 1$

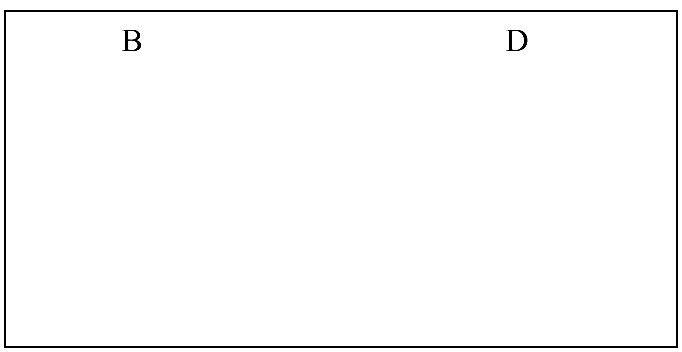
We note that $B \cap D = \{\ \}$, abd $n(B \cap D) = 0$

Therefore, $P(B) = \frac{3}{6}$ P(D)

$= \frac{1}{6}$ $P(B \cap D) = \frac{0}{6} = 0$

In this case B and D cannot occur together.

We can now draw the following Venn Diagram



To avoid the errors involved in rounding off, probabilities have been given as fractions…..

We note that $\frac{3}{6} + 0 + \frac{1}{6} + \frac{4}{6} \neq 1!$

The remaining region therefore must have a probability of $1 - \frac{4}{6} = \frac{2}{6}$.

Compare this with the number of elements of S that are not in B or D and their corresponding proportion to n(S). What do you notice?

***A*CTIVITY: MT/2/3-3**

Draw a Venn diagram to represent the following probabilities and in each case check that P(S) = 1.

(i)	$P(A) = \frac{3}{4}$	$P(B) = \frac{5}{8}$	$P(A \cap B) = \frac{1}{2}$
(ii)	$P(A) = \frac{1}{2}$	$P(B) = \frac{1}{3}$	$P(A \cap B) = \frac{1}{4}$
(iii)	$P(A) = 0.35$	$P(A \cup B) = 0.65$	$P(A \cap B) = 0$

✓ Check your answers with those given at the end of this unit.

(d) Solving Probability Problems using set theory

Now that you have learnt to represent probabilities in Venn diagrams, we shall use this knowledge to solve problems involving probability, which can be represented in venn diagrams.

In all cases, it will be important for you to recall the following:

1. $P(S) = 1$, where S is the sample space.

2. $P(A') = 1 - P(A)$, where A' is the event that A does not occur. (i.e. $A' =$ not A).

3. $P(A \cap B)$ is the probability that A and B occur together or simultaneously.

4. $P(A \cup B)$ is the probability that event A or event B or both occur.

5. A given region in the Venn diagram represents an event, which event we refer to in the same way as we would refer to sets; for example:

- The region for the intersection represents the event $A \cap B$.
- The region outside the circles but within the rectangle represents the event $A' \cap B'$, which is not A and not B occur.

Let us have some example of solving probability problems using set theory.

Example 1

Two events A and B are such that:

$P(A') = 0.65$, $P(A \cup B) = 0.65$ and $P(A \cap B) = 0$

If P(B) = P, find the value of P. Also find the value of $P(A' \cap B')$.

Solution

From the given information, we can find p(A) as follows:

$$
\begin{aligned}
P(A) &= 1 - P(A') \\
&= 1 - 0.65 \\
&= 0.35
\end{aligned}
$$

We are also given that $P(A \cup B) = 0.65$;

And from set theory, we know that $n(A \cup B) = n(A) + n(B) - n(A \cap B)$,
Which, when divided through by n(S) gives.

$$\frac{n(A \cup B)}{n(S)} = \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} - \frac{n(A \cup B)}{n(S)}$$

This is equivalent to:

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

Therefore $0.65 = 0.35 + P - 0$ and $P = 0.30$

Further more, if the region $A' \cap B'$ is called E.
Then $0.35 + 0 + 0.30 + E = 1$, giving E as 0.35.

We then draw a Venn diagram and fill in the probabilities for A, $A \cap B$, B and E as below:

*

Note: The formula

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

Is referred to as the general addition rule of probabilities.

Example 2

Given two events A and B such that:

$P(A) = \frac{1}{2} \quad P(B) = \frac{1}{3}$ and $P(A \cap B) = \frac{1}{4}$

Use a Venn diagram to find $P(A'UB)$ and $P(A' \cap B)$

Solution

We draw the Venn diagram and fill in the probabilities beginning from the centre (intersection) outwards.

*

Here is the working for obtaining the probabilities.

$$a = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

$$b = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

$$c = 1 - \left(\frac{1}{4} + \frac{1}{4} + \frac{1}{12} -\right) = \frac{5}{12}$$

Go through the working and be sure that you understand the working.
Remember, we are using set theory, but working with probabilities.

Can you now identify the regions (s) representing $A' \cup B$? These regions give us

$$P(A' \cup B) = \frac{1}{4} + \frac{1}{12} + \frac{5}{12} = \frac{9}{12} = \frac{3}{4}$$

Similarly, $P(A' \cap B) = \frac{1}{12}$

☞ **Note:** $(A' \cup B) = P(A') + P(B) - P(A'B)$. Use the values above to verify this.

Example 3

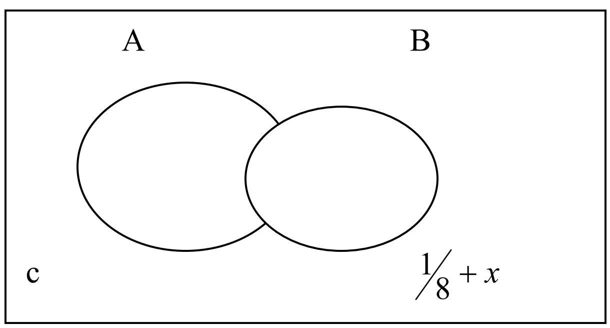
For events A and B such that $P(A) = \frac{1}{2}$

$P(B') = \frac{5}{8}$ and $P(AUB) = \frac{3}{4}$

Find $P(A \cap B)$, $P(A' \cap B')$ and $P(A' \cup B)$

Solution

We shall draw the venn diagram and fill in probabilities beginning from the centre. Since $P(A \cap B)$ is not given, let $P(A \cap B)$ be x. The other regions will relate to x as shown below.



We obtain the probabilities in the regions of the Venn diagrams working as follows:

Since $P(A) = \frac{1}{2}$
then $a + x = \frac{1}{2}$
therefore $a = \frac{1}{2} - x$

From $P(B') = \frac{5}{8}$

$P(B) = 1 - \frac{5}{8}$

$$= \frac{3}{8}$$

$$\text{Then } b + x = \frac{3}{8}$$

$$\text{and } b = \frac{3}{8} - x$$

$$\text{also } P(S) = 1$$

$$\text{Hence } (½ - x) + x + (\frac{3}{8} - x) + c = 1$$

$$\frac{7}{8} - x + c = 1$$

$$\text{Therefore } c = \frac{1}{8} + x$$

But remember we are given that $P(AUB) = \frac{3}{4}$

Which means that $\frac{1}{2} - x + x + \frac{3}{8} - x = \frac{3}{4}$

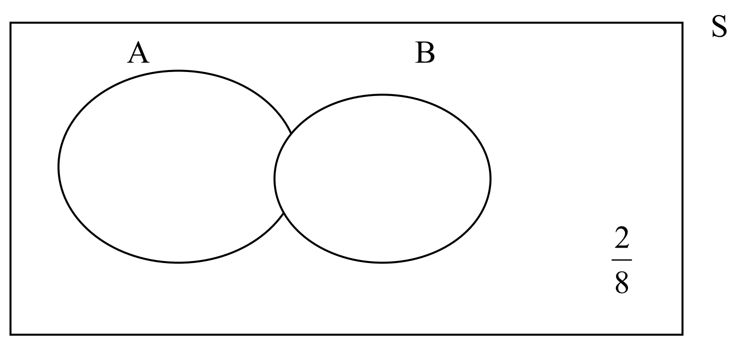
The expression on the LHS represents the values of the probabilities in the region that represents $A \cup B$.

$$\text{Hence } \frac{7}{8} - x = \frac{3}{4}$$

$$\frac{7}{8} - \frac{3}{4} = x$$

$$\text{Therefore } x = \frac{1}{8}$$

Drawing another Venn diagram, with x substituted for with $\frac{1}{8}$, we have



We can then write down the answers as:

$P(A \cap B) = \frac{1}{8}$

$P(A' \cap B') = \frac{2}{8}$

$P(A' \cup B) = \frac{5}{8}$

To get a clearer picture of how to obtain any of these probabilities, but especially the last one, you can shade the corresponding regions and note the probabilities in them.

Also, note that from the general addition rule of probabilities,
$P(A' \cup B) = P(A') + P(B) - P(A' \cap B)$
Now go through the examples one more time before attempting Activity MT/2/3-4.

***ACTIVITY:* MT/2/3/-4**

1. Events A and B are such that $P(A') = \frac{5}{8}$, $P(A \cup B) = \frac{7}{8}$ and $P(A \cap B) = \frac{1}{4}$

   Find $P(A)$, $P(B)$, $P(A \cap B')$, $P(A' \cap B')$
   $P(A' \cap B)$, $P(A' \cup B)$, $P(B' \cup A)$ and $P(A' \cup B')$
   What is the relationship between $P(A \cap B)$ and $P(A' \cup B')$?

2. Find the values of $P(A')$, $P(B')$, $P(A \cup B)$, $P(A \cup' B)$ and $P(A' \cap B)$ given that:
   $P(A) = \frac{3}{4}$, $P(B) = \frac{5}{8}$ and $P(A \cap B) = \frac{1}{2}$

3. Given that $P(A) = \frac{1}{4}$, $P(B) = \frac{2}{5}$ and $P(A \cup B) = \frac{1}{2}$

   Find $P(A \cap B)$, and $P(A' \cap B')$

✓ Check your answers with those given at the end of this unit.

(e) **Sample Space Diagrams**

So far, you have learnt how to list sample spaces as sets using curly brackets, and representing them in Venn diagrams. Sometimes it may be easier to give the sample space in form of a diagram, similar to a coordinate system.

As you very well know, we commonly use a two coordinate system, with a horizontal axis (x – axis) and a vertical axis (y – axis). Even for sample space diagrams, we shall use horizontal and vertical axes to represent different outcomes of given trials.

Sample space diagrams are most useful when we have combined events or when the number of elements in S is big. A single event in a sample space diagram is also called a sample point. In this section, we shall build sample diagrams from the simple sample spaces we have been listing to bigger sample spaces involving more trials. Study and work through the following examples.

Example 1

When a coin is tossed:

$$S = \{H, T\},$$

When H is a head and T is a tail.

We shall use this sample space to get other sample spaces below:

Example 2

If two coins are tossed, each can show a head or a tail. (This trial is similar to tossing one coin twice).

We have the following sample space diagram.

1st coin

2nd coin

	H	T
H	HH	HT
T	TH	TT

If you study the diagram, you will notice that there are four sample points, HH, HT, TH and TT. Each point is at the intersection of a possible horizontal outcome and a possible vertical outcome.

Example 3

If three coins are tossed, we have the following diagram.

Outcomes for tossing 2 coins

3rd coin

	HH	HT	TH	TT
H	HHH	HHT	HTH	HTT
T	THH	THT	TTH	TTT

How many sample points are in this sample space diagram?

In this way, as you may have noticed, we easily get all the possible outcomes or sample points.

Example 4

Building on from the last 3 examples, we have the following diagram for tossing four coins.

Outcomes for tossing 3 coins

4th coin

	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
H	HHHH	HHT	HHTH	HHTT	HTHH	HTHT	HTTH	HTTT
T	THHH	THHT	THTH	THTT	TTHH	TTHT	TTTH	TTTT

How many sample points are in this diagram?

Draw a table summarizing the number of coins tossed and the number of sample points in the sample space diagram. Do you think you can use your summary table to predict the sample points when 5 coins are tossed? Tyr to make a formula that can predict the sample points for tossing n coins.

I hope you have enjoyed building sample space diagrams for tossing coins. Now try Activity MT/2/3-5

Good Luck1

ACTIVITY MT/2/3-5

1. Suppose we let H = 1 and T = 0, write out the sample space diagrams beginning with tossing 1 coin to 4 coins using 1s and 0s. If each sample point is taken to represent a number in binary, find the decimal equivalent for each sample point in each sample space diagram. Find also the relationship between 2 and n(S) in each case.

2. Build sample space diagrams for tossing 5 and 6 coins respectively.

✓ Check your answers with those given at the end of this unit.

Example 5 of a sample space diagram

When a die is tossed, S = {1, 2, 3, 4, 5, 6}.
If two dice are tossed, each can show any of the six equally likely possible outcomes.

The sample diagram for two dice then looks like this.

1st die

2nd die

	1	2	3	4	5	6
1	1,1	1,2	1,3	1,4	1,5	1,6
2	2,1	2,2	2,3	2,4	2,5	2,6
3	3,1	3,2	3,3	3,4	3,5	3,6
4	4,1	4,2	4,3	4,4	4,5	4,6
5	5,1	5,2	5,3	5,4	5,5	5,6
6	6,1	6,2	6,3	6,4	6,5	6,6

How many sample points do we have altogether? Most times, we are interested in the sum of the outcomes of the two dice. In this case, our sample points are the sums as shown in the following diagram.

1st die

2nd die

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

What are the possible sums?
How many sample points are there?

Write down in tabular form each possible sum and the probability of getting it.

We could also be interested in the product of the outcomes of the two dice and come up with the following sample space diagram.

1st die

2nd die

•	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

What are the possible products? How many sample points are there?

Make a table showing each possible product and the probability of getting it. Now try drawing a sample space diagram for the absolute difference of the outcomes of the two dice (that is the difference irrespective of whether it is negative).

Example 6

Before beginning a game of chess, the players toss a die and a coin to decide on who take which colour of the pieces. The possible outcomes of the tosses can be represented in the sample space diagram below.

Die

Coin

	1	2	3	4	5	6
H	HI	H2	H3	H4	H5	H6
T	T1	T2	T3	T4	T5	T6

Example 7

The Headmistress of St. Catherine's Secondary School asked the Game's Master to group the students according to the colour of their games' T-shirt and their favourite game or sport. The colours to be chosen from were Red (R), Blue (B), Green (G), White (W) and Yellow (Y). Whereas the games or sports were Football (F), Netball (N), Volleyball (V), Hockey (H) and Cricket (C).

To ensure that he had all the possible outcomes in an orderly way, the Game's Master drew the following sample space diagram.

Colour of Game's T-Shirt

Game or Sport

	R	B	G	W	Y
C	CR	CB	CG	CW	CY
F	FR	FB	FG	FW	FY
H	HR	HB	HG	HW	HY
N	NR	NB	NG	NW	NY
V	VR	VB	VG	VW	VY

From the diagram, the Game's Master could easily see that he would have 25 different groups. This could help him organize the activities for the season.

TOPIC 2: MUTUALLY EXCLUSIVE AND INDEPENDENT EVENTS

In this topic you are going to learn about:

- Mutually exclusive events.
- Independent events.

(a) Mutually Exclusive Events

When learning about sets, you came across what are called disjoint sets; that is sets which do not have any element in common.

For example if $A = \{a, b, c\}$ and
$B = \{e, f, g, h\}$

Then A and B are disjoint sets and $A \cap B = \{\ \}$; an empty set.

Definitely, $n(A \cap B) = 0$

If we now consider A and B as events, we can say A is the event that one of the first three elements of the alphabet is chosen. Similarly B is the event that one of the letters from e to h inclusive, is chosen.

Clearly $A \cap B$ us an empty set. In other words, A and B cannot occur together. (We cannot have a letter in set A and at the same time it is B).

A and B are said to be MUTUALLY EXCLUSIVE events.

Consequently, since $n(A \cap B) = 0$,

Then $\boxed{P(A \cap B) = \frac{n(A \cap B)}{n(S)} = 0}$ whenever A and B are mutually exclusive events.

Examples of Mutually Exclusive Events

1. A: Adyeri is in P.1
   B: Adyeri is in P.4

2. A: Today is a Monday
   B: Today is Wednesday

3. A: An integer chosen at random is even.
   D: An integer chosen at random is odd.

4. H: The Headmaster is in his office.
   E: The Headmaster has gone to the District Headquarters.

5. M: Mother has gone to the well to collect water.
   W: Mother is weeding the garden.

ACTIVITY: MT/2/3-6

1. Write down ten different example of mutually exclusive events.

2. Given that A and B are two mutually exclusive events with
   $P(A') = 0.55$ and $P(B) = 0.25$, find $P(A \cup B)$.

3. What is the probability that Kato, a DEPE student goes to Kibuli PTC for his face-to-face sessions and he also goes to Kabwangasi PTC for the same face-to-face session? Give a reason for your answer.

4. An integer is chosen from the set of all integers and B is the event that the chosen number is positive, while N is the event that the chosen number is negative. What is $n(B \cap N)$ and $P(B \cap N)$? What can you say about the two events B and N?

5. What is the probability that Acan is a pupil in P.3B and Acan is the class teacher of P.3B? Which are the two mutually exclusive events in this case?

✓ Check your answers with those given at the end of this unit.

(b) **Independent Events**

Two events A and B are said to be independent if the knowledge that A has occurred, will occur or can occur does not affect the probability that B will occur.

For any two such events A and B, $P(A \cap B) = P(A) . P(B)$

Examples of Independent Events

1. A: a coin is tossed once and lands heads up.
   B: A coin is tossed a second time and lands tails up.

2. A: My class teacher in P.1 was a female.
   B: My class teacher in S.1 was a female

3. A: The first throw of a die in a six.
   B: The second throw of a die is a six.

4. A: When Amooti draws a card from a well shuffled pack of 52 playing cards, it is found to be an ‘ace’ (A).
   B: When Amooti tosses a coin, it lands heads up.

5. A: May’s first born child is a boy.
   B: Mary’s second born child is a girl.

When that for the above examples;

In Example 1:

$$S = \{H, T\}$$

$$A = \{H\}, \quad P(A) = \frac{1}{2}$$

$$\text{Therefore } P(A \cap B) = P(A), \; P(B)$$

$$= \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

In Example 2:

$$S = \{\text{Male, Female}\}$$

$$A = \{\text{Female}\}, \quad P(A) = ½$$

$$B = \{\text{Female}\}; \quad P(B) = ½$$

$$\text{Therefore } P(A \cap) = ½ \times ½ = ¼$$

In Example 3:

$$S = \{1, 2, 3, 4, 5, 6\}$$

$$A = \{6\}, \quad P(B) = \frac{1}{6}$$

$$\text{Therefore } P(A \cap) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

In example 4:

$$n(S) = 52$$

$$N(A) = 4; \text{ so that } P(A) = \frac{4}{52}$$

$$P(B) = \frac{1}{2}$$

$$\text{Therefore } P(A \cap B) = \frac{4}{52} \times \frac{1}{2}$$

$$= \frac{1}{26}$$

and in example 5: $P(A) = ½$
$P(B) = ½$

and $P(A \cap B) = ½ \times ½$

$= ¼$

***A*CTIVITY: MT/2/3-7**

1. Write down five different examples of independent events. For each example, find the probability $P(A \cap B)$, that is the probability that the two events occur together.

2. The probability that a student chosen at random names Mathematics as their favourite subject is 0.70, Find the probability that two unrelated students will name Mathematics as their favourite subject.

3. The probability that Mrs. Takembo will be elected the next L.C 1 Chairperson of Namakwekwe South village is 0.65. Find the probability that two different voters from two different polling stations vote for Takembo.

✓ Check your answers with those given at the end of this unit.

TOPIC 3: PROBABILITY TREES AND THEIR USE IN SOLVING PROBLEMS

In this topic you are going to learn about:

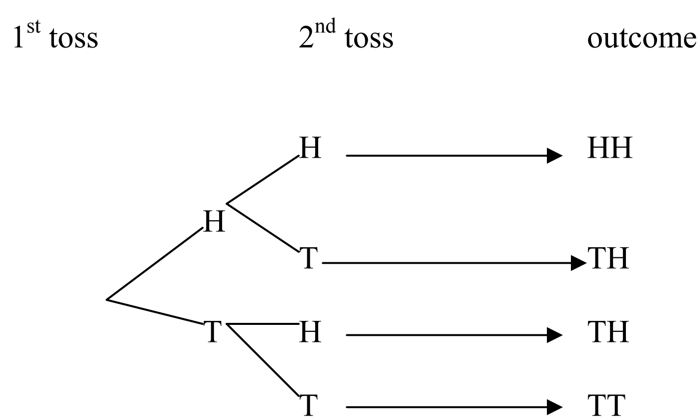
- Tree diagrams.
- Probability trees.
- Using probability trees to solve problems.

(a) TREE DIAGRAMS

You have already learnt how to list a sample space for fairly simple situations. Sometimes, however, it demands a lot of patience to produce a complete sample space. We shall look at a different way in which we can arrange a sample space using a tree diagram. We can define a tree diagram as a sample space that has a starting point and at least two branches originating from this starting point. At the end of each branch, we can have more branches.

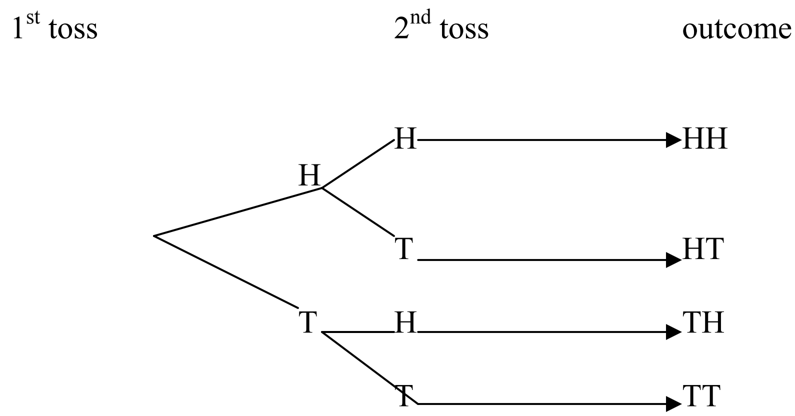
Examples

1. When a coin is tossed twice, we have the sample space S = {HH, HT, TH, TT}. Using a tree diagram, we can arrange this sample space in the following way.



We call this a tree diagram.

Let us examine this tree diagram more carefully.



In the regions of the diagram under 1st toss, we note that when a coin is tossed, it can land either Head (H) or Tail (T) up.

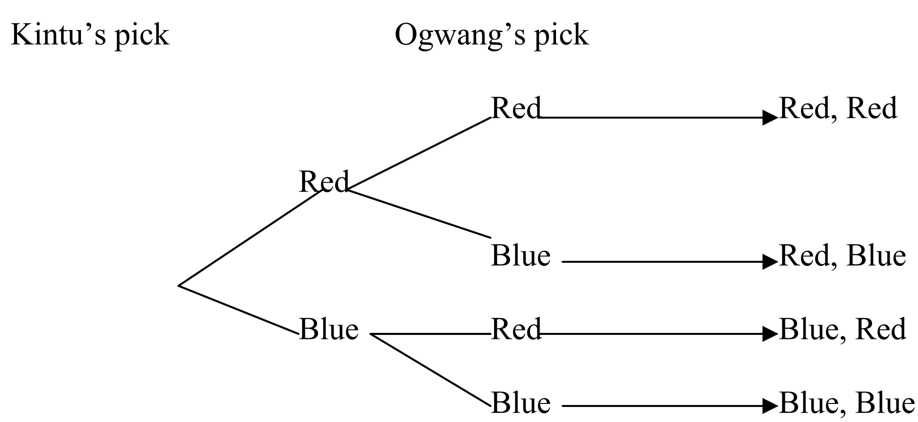
Therefore, from the starting point, we have two branches, each representing a possible outcome of a single toss of the coin.

You may recall the idea of independent events. The outcomes of tossing the coin a second time are independent of the outcomes of the first toss. Whatever the outcome of the first toss, therefore, we still expect a Head (H) or a Tail (T) on the second toss.

This is well represented in the region of the tree diagram under 2nd toss, where for each outcome of the first toss, you have two branches each representing a possible outcome. In the end, we can say that you have four complete branches, each representing a possible outcome of two tosses of a coin.

The top branch as you can see from the diagram is HH; that is both tosses show heads. In the same way, we can identify all the other outcomes that make up the sample space.

2. A bag has some red and some blue balls. Kintu picks a ball and goes out to play with it. Ogwang also comes and picks a ball and goes out to play with it. Represent this information on a tree diagram.

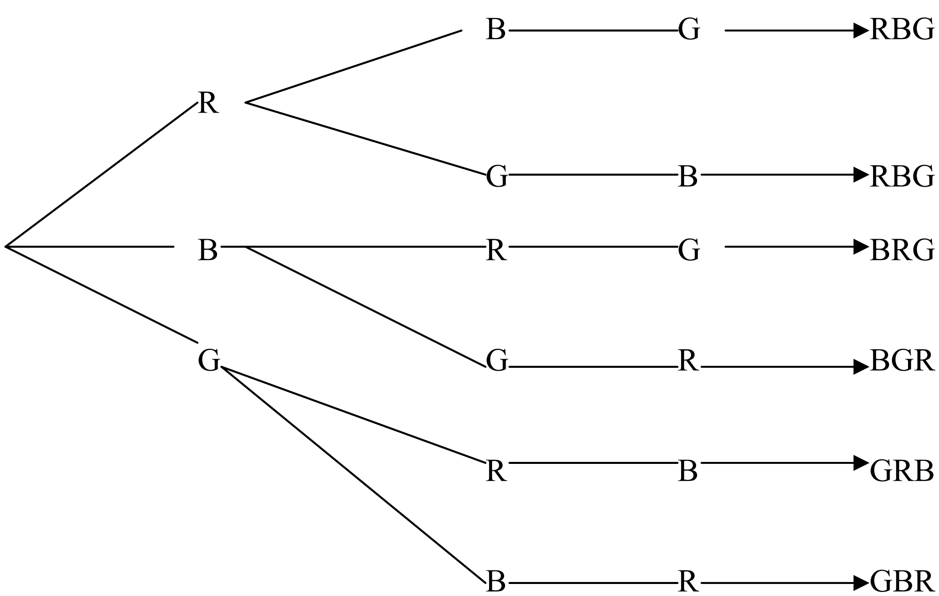


From this tree diagram we can see that Kimtu could have picked either a red or a blue ball. Whichever one he picked, Ogwang could also have picked a red or a blue ball and the four complete branches represent all the possible outcomes of their picks. Both could have picked red balls, blue balls or each could have picked a different colour.

3. A bag contains a red (R) a blue (B) and a green (G) ball of the same size. A ball is taken out and put on the desk. A second ball is taken out and also put on the desk. Then the last ball is taken out as well. Draw a tree diagram of the six possible outcomes.

   In this example, it is important to note that initially, we can take out any of the three balls. Next we have two balls to choose from then finally, there will be only one ball.

1st pick 2nd Last pick



How does the second pick depend on the first pick? Remember there are three balls, so if the first pick was a red ball, the second can only be blue or green. If the first was red, the second green, then the third has to be blue. This is how the six possible outcomes arte got.

ACTIVITY: MT/2/3-8

Draw tree diagrams to represent the following:

(i) A two digit number is formed as follows:
The tens digit is chosen from (3, 4, 5) and the ones digit is chosen from (6, 8,2).

(ii) A two digit number is formed by choosing both digits from (3, 4, 8, 9). However, the same digit may not be chosen twice.

(iii) Repeat (ii) if the same digit may be used twice.

✓ Check your answers with those given a the end of this unit.

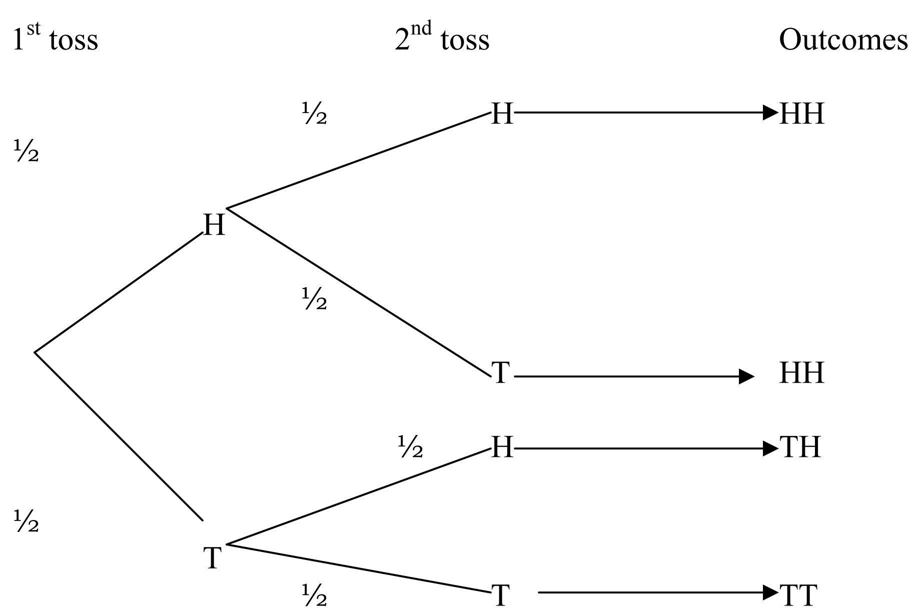
(b) **Probability Tree**

We have been drawing tree diagrams showing possible outcomes of given trials without considering the probabilities with which those outcomes occur. If these probabilities are indicated on their respective branches, our tree diagrams become probability trees. Sometimes, tree diagrams actually refer to probability trees.

Examples

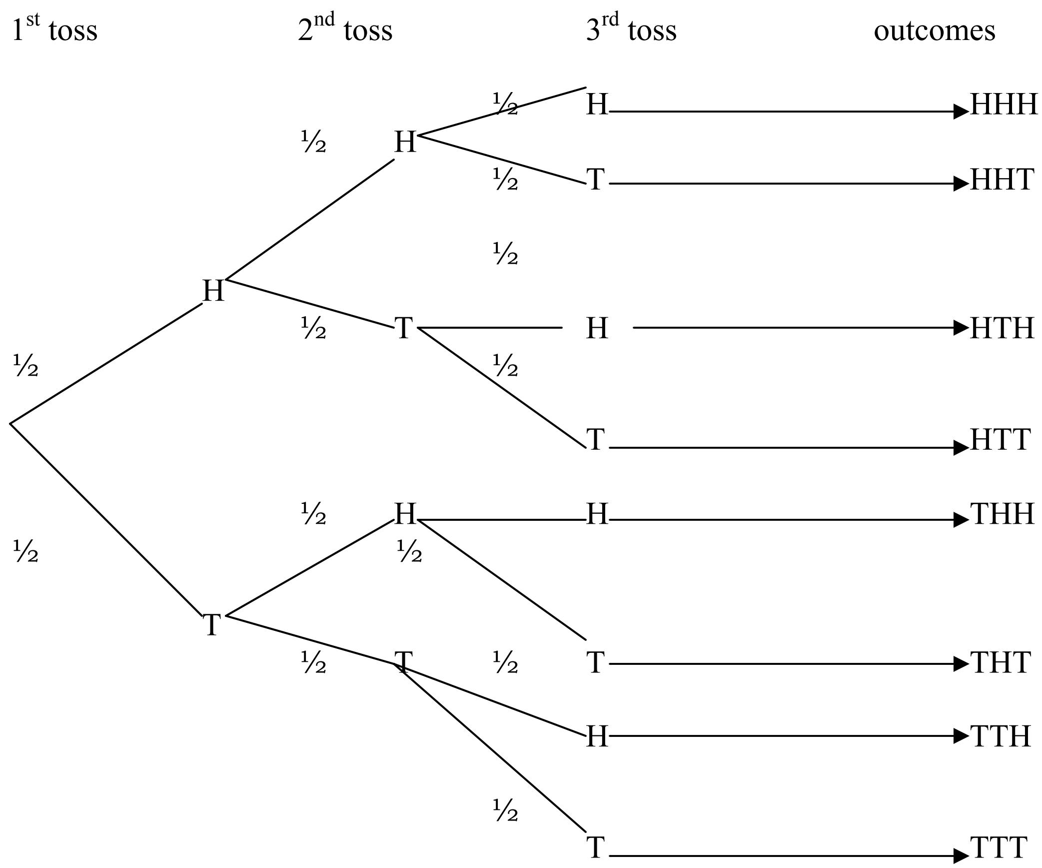
1. When a coin is tossed twice, what are the possible outcomes of each toss? What are their respective probabilities?

   We shall draw a probability tree to show us the required information.

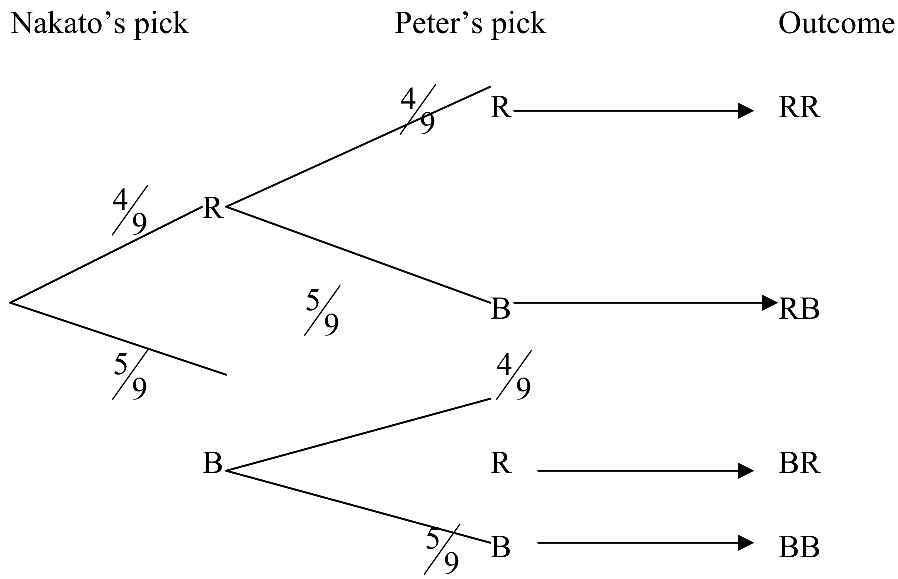


Whenever a coin is tossed, the probability of a head is ½ and the probability of a tail is also ½.

2. Using a tree diagram, write downs all the possible outcomes that can occur when a coin is tossed three times. Indicate the respective probabilities for each toss.

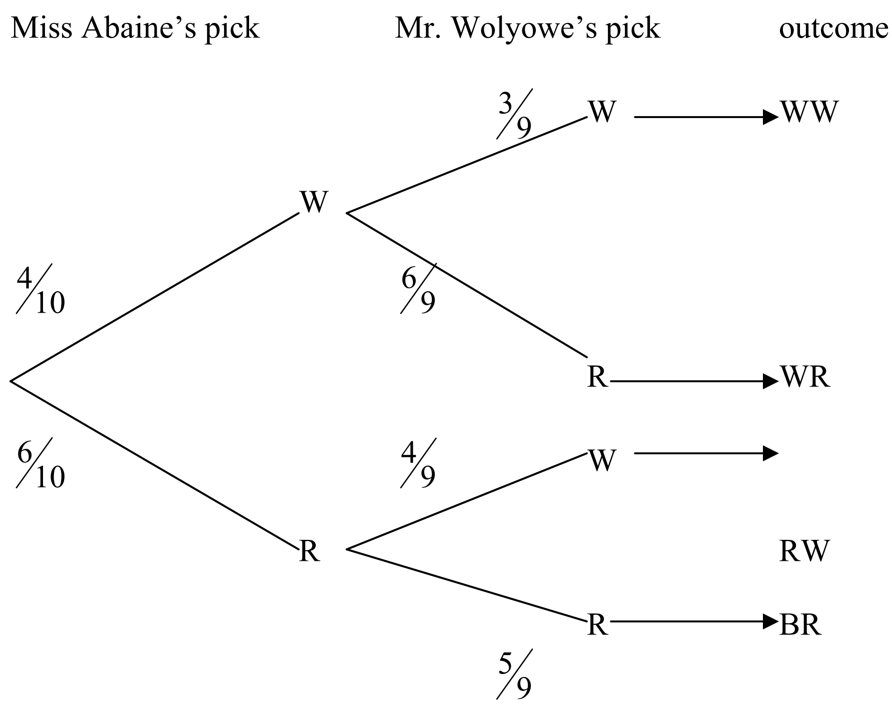


3. A bag contains 4 red (R) and 5 blue (B) balls. Nakato picks a ball and goes out to play with it. When she is tired, Peter asks her for it. She refused and brings it back into the bag. Peter then comes and picks a ball and goes out to play with it. Draw a tree diagram for the possible picks and indicate the probability of each pick.



4. A box contain 4 white (W) and 6 red (R) pieces of chalk. Miss Abaine picks a piece of chalk and goes to teach P.6.B. Mr. Wolyowe picks a second piece of chalk and goes to teach P.2C. represent this information on a probability tree.

It is important to note that when Mr. Wolyowa picks the second piece of chalk, the total number of pieces will have reduced by the one which Miss Abaine picked. Also, we do not know which colours the two teachers pick.



From the above probability tree, we note the following:

(i) If Abaine picked white, Wolyawe would have three white and six red to pick from.

(ii) If Abaine picked red, Wolyawe would have four white and five red to pick from.

(iii) The probabilities on all the branches starting from the same point always add up to 1. This is true in all cases of probability trees.

ACTIVITY: MT/2/3-9

1. A tin contains two pink and three sweets. A child selects one, looks at it and puts it back. After shaking the tin, she selects a second sweet. Represent this information on probability tree.

2. Repeat question 1 for the case when the child eats the first sweet and then selects a second one.

3. In a students trunk there are two pairs of shoes. One pair is light brown and the other pair is dark brown. Otherwise the shoes are very much alike. One night when thee are no lights, the student reaches into the trunk and takes out two shoes at random.

   Draw a probability tree for his possible outcomes. Assume that the student takes one shoe at a time but does not replace the first one before picking the second.

✓ Check your answers with those given at the end of this unit.

(c) USING PROBABILITY TREES TO SOLVE PROBLEMS

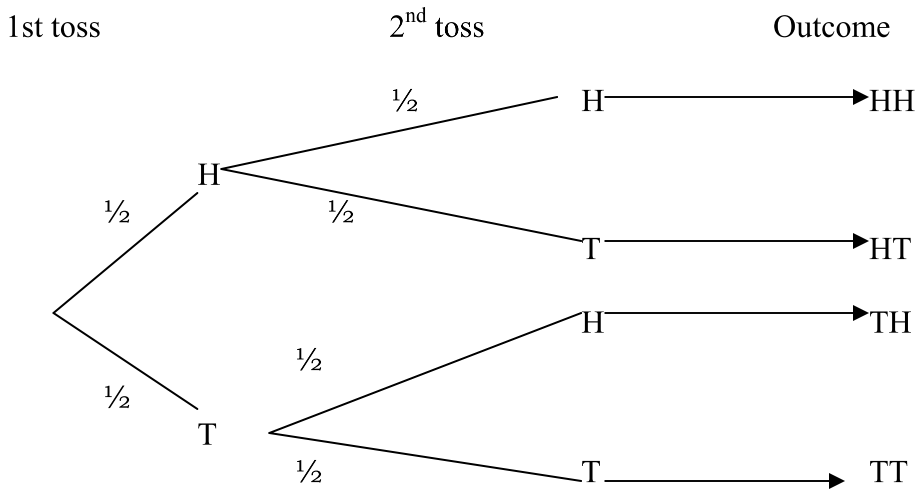
Once we can tell which outcome is represented by a certain route on the probability tree, then we would be in position to quickly solve any related problems. All we need to recall is that in probability we interpret the "and" as multiplication and the "or" as addition. We only need to be careful about using the correct probability for each trial's outcome, bearing in mind that probabilities on all branches from the same starting point always add up to 1.

Remember the tree diagram is another form of representing a sample space. It must therefore clearly indicate all the possible outcomes of a given trial.

Examples

1. Use a probability tree to find the probability that in two tosses of a coin.

   (i) A head occurs twice.
   (ii) A tail is obtained once.



(i) P(head and head) $=$ P(HH)

$= \frac{1}{2} \times \frac{1}{2}$

$= \frac{1}{4}$

(ii) P(Tail once) $=$ P(Head and tail or tail and Head)

$=$ P(HT or TH) $=$ P(HT) + P(TH)

$= (\frac{1}{2} \times \frac{1}{2}) + (\frac{1}{2} \times \frac{1}{2}) = \frac{1}{2}$

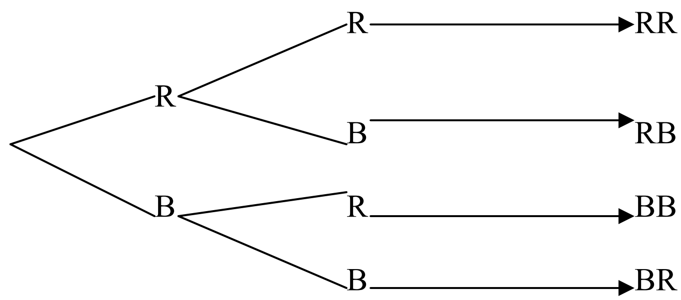
2. A bag contains 4 red (R) and 5 blue (B) balls. Nakato picks a ball and goes out to play with it. When she is tired, Peter asks her for it. She refuses and brings it back into the bag. Peter then comes and picks a ball and goes out to play with it.

   Find the probabilities that both Peter and Nakato pick.

   (i) Red balls
   (ii) Blue balls
   (iii) Balls of different colours.
   (iv) Balls of the same colour.

Let us have the probability tree for this example.

Nakato's pick   Peter's pick   Outcome



(i) P(both red balls) = P(RR) = $\frac{4}{9} \times \frac{4}{9} = \frac{16}{81}$

(ii) P(both blue balls) = P(BB) = $\frac{5}{9} \times \frac{5}{9} = \frac{25}{81}$

(iii) P(balls of different colours) = P(red and blue or blue and red)

$$\frac{(4}{9} \times \frac{5)}{9} + \frac{(5}{9} \times \frac{4)}{9} = \frac{40}{81}$$

(iv) P(balls of the same colour) = P(both red or both blue)

$$= \frac{(4}{9} x \frac{4)}{9} + \frac{(5}{9} \times \frac{5)}{9}$$

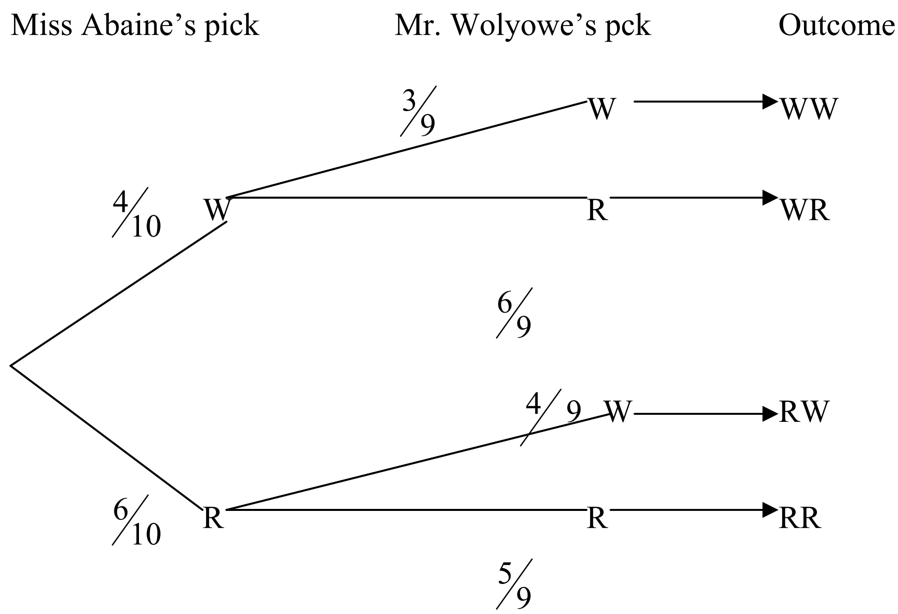
$$= \frac{41}{81}$$

3. A box contain 4 white (W) and 6 red (R) pieces of chalk. Miss Abaine picks a piece of chalk and goes to teach P.6.B. Mr. Wolyowe picks a second piece of chalk and goes to teach P.2C. Represent this information on a probability tree.

(a) Find the probabilities that both teachers pick:

(i) White chalk
(ii) Red chalk
(iii)The same colour of chalk
(iv)Different colours of chalk

(b) Find the probability that Miss Abaine picks a white piece of chalk and Mr. Wolyawe picks a red piece of chalk.



(a) (i) P(both white) = P(WW) $= \frac{4}{10} \times \frac{3}{9} = \frac{4}{30} = \frac{2}{15}$

(ii) P(both red) = P(RR) $= \frac{6}{10} \times \frac{5}{9} = \frac{1}{3}$

P(same colour) = P(WW or RR) $= \frac{2}{15} + \frac{1}{3} = \frac{7}{15}$

P(different colours) = P(WR or RW) $= \frac{4}{10} \times \frac{6}{9} + \frac{6}{10} \times \frac{4}{9} = \frac{8}{15}$

(b) P(W then R) $= \frac{4}{10} \times \frac{6}{9} = \frac{4}{15}$

***A*CTIVITY: MT/2/3-10**

Solve the following problems using probability trees.

1. A tin contains 7 red and 5 blue T-shirts. Two T-shirts are picked at random. Find the probability of picking.

   (i) 1 red and 1 blue T-shirt.
   (ii) T-shirts of the same colour assuming that the first T-shirt is not replaced before the second one is picked.

2. Three people are to be chosen to form a committee, from 3 women, 2 youths and 4 men. Find the probability that:

   (a) The committee has only women.
   (b) All the youths are on the committee.
   (c) There is only one man on the committee.

3. A bag has 5 sweets; 2 with hard centres (H) and 3 with soft centres (S). Maria chooses a sweet and eats it. Kafeero also chooses a sweet and eat it. Find the probability that:

   (a) Both choose sweets with soft centres.
   (b) Maria chooses a sweet with a hard centre and Kafeero chooses one with a soft centre.

4. Cards marked with the numbers 1, 2, 3, 4, 5 are put in a bag and two are drawn at random, the first one being replaced before the second card is drawn. Find the probability that:

   (a) Both have odd numbers on them.
   (b) Both have even numbers on them.
   (c) Both have the same number.

5. A coin is tossed three times. Find the probability of getting;

   (a) All heads
   (b) All tails
   (c) At least one head and one tail;

✓ Check your answer with those at the end of this unit.

SUMMARY OF UNIT 3

In this unit, you have learnt about solving probability problems using set theory and probability trees. You have also been introduced to mutually exclusive and independent events and probability distributions. Now you are in position to use a variety of approaches to solving probability problems.

Congratulations!

GLOSSARY

Independent events: events such that occurrence of one does not affect the probability that the other(s) occurs.

Mutually exclusive events: events that cannot occur together.

Probability tree: representation of a sample space using branches from a starting point.

NOTES AND ANSWERS TO ACTIVITIES

ACTIVITY: MT/2/3-1

Any 5 acceptable, well defined sample spaces e.g.

S	=	letters of the alphabet
S	=	positive integers
S	=	real numbers
S	=	all animals
S	=	all domestic birds

ACTIVITY: MT/2/3-2

Any well defined subsets of S. e.g.

(i) A = (a, e, I, o, u), B = (b, d, h, k, l, t) C = (x, y, z)

(ii) K = (1, 2, 3,….,10) m M = (100, 200,300,…,900)
L = (2, 4, 6,8,…)

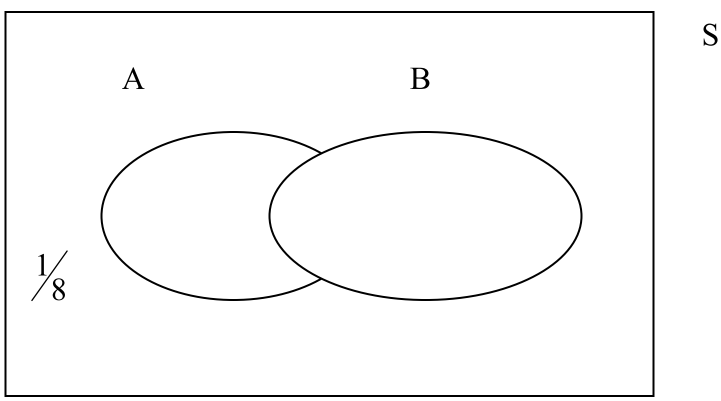
(iii) A = (½, 0, -½) B = (0, 1, 0, 2,…..0.9)
C = (-4, -3, -2, -1, 0, 1, 2, 3, 4)

(iv) A = (wild animals), B = (domestic animals)
C = (monkeys)

(iv) A = (chicken), B = (ducks) C = (turkeys)

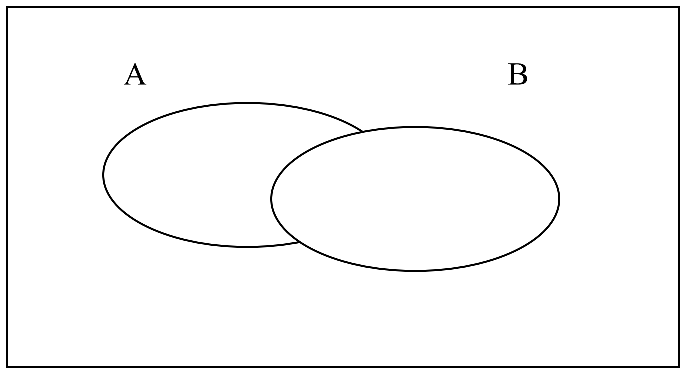
ACTIVITY: MT/2/3-3

(i)



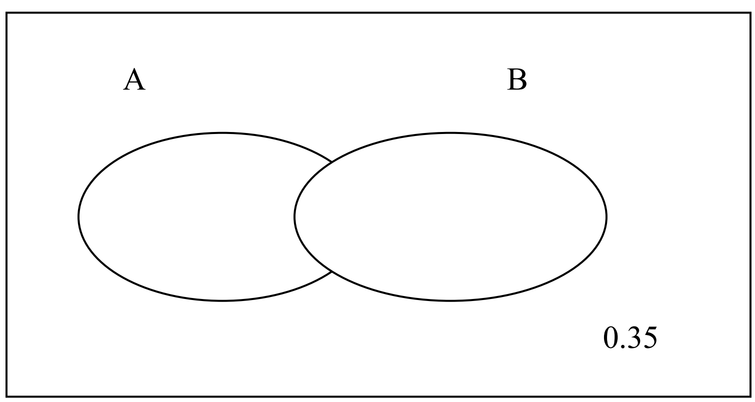
$$\frac{1}{4} + \frac{1}{2} + \frac{1}{8} + \frac{1}{8} = \frac{2 + 4 + 1 + 1}{8} = 1$$

(ii)



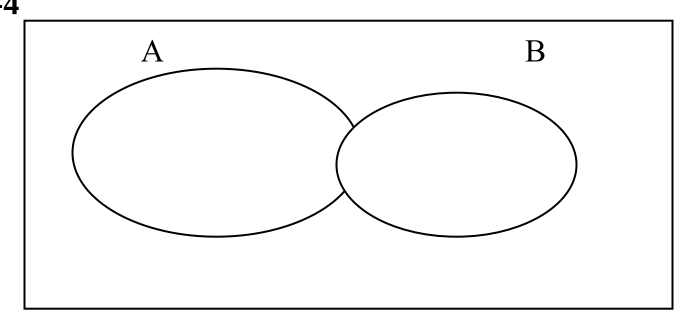
$$\frac{1}{4} + \frac{1}{4} + \frac{1}{12} + \frac{5}{12} = \frac{3+3+1+5=1}{12}$$

(iii)



$$0.35 + 0 + 0.30 + 0.35 = 1$$

ACTIVITY: MT/2/3-4



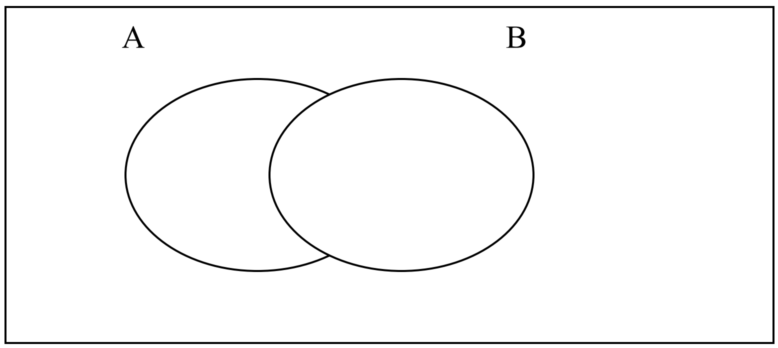
$P(A) = \frac{3}{8}$, $P(B) = \frac{3}{4}$; $P(A \cap B') = \frac{1}{8}$;

$P(A' \cap B') = \frac{1}{8}$; $P(A' \cap B) = \frac{1}{2}$; $P(A' \cup B) = \frac{2}{8}$;

$P(B' \cup A) = \frac{1}{2}$; $P(A' \cup B') = \frac{3}{4}$

$P(A' \cup B') = 1 - P(A \cap B)$

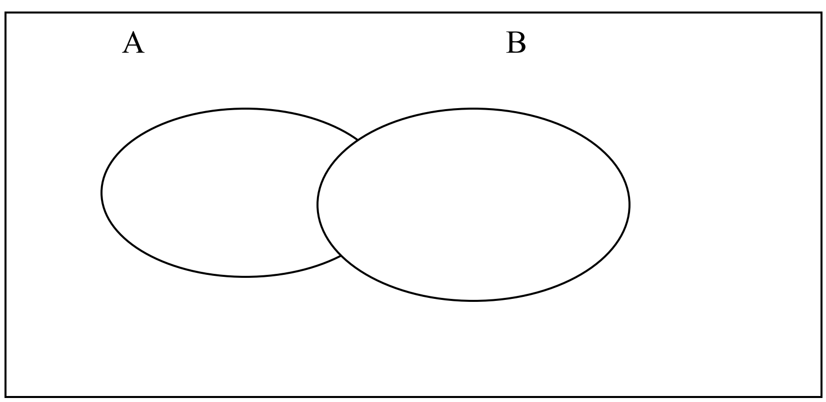
2.



$P(A') = \frac{1}{4}$; $P(B') = \frac{3}{8}$; $P(A \cup B) = \frac{7}{8}$;

$P(A' \cup B') = \frac{1}{2}$; $P(A' \cap B) = \frac{1}{8}$

3.



$P(A \cap B) = \frac{3}{20}$; $P(A \cap B') = \frac{1}{10}$; $P(A' \cap B') = \frac{1}{2}$

ACTIVITY: MT/2/3-5

1. For 1 coin, S = (1,0)

   2 coins, S = (00,01, 10, 11)

   3 coins, S = (000,0001,010,011,100,101,110,111)

   4 coins, S = (0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111, 1000, 1001, 1010, 1011, 1100, 1101, 1111)

Decimal equivalent are:

S = (1,0)
S = (0,1,2,3), (0,1,2,3,4,5,6,7)
S = (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15)

In each case, n(S) = 2^k, where k is the number of coins being tossed.

2. For 5 coins 5th coin

Outcomes for
Tossing 4 coins

	H	T
HHHH	HHHHH	HHHHT
HHHT	HHHTH	HHHTT
HHTH	HHTHH	HHTHT
HHTT	HHTTH	HHTTT
HTHH	HTHHH	HTHHT
HTHT	HTHTH	HTHTT
HTTH	HTTHH	HTTT
HTTT	HTTTH	THHHT
THHH	THHHH	THHTT
THHT	THHTH	THTHT
THTT	THTTH	THTTT
TTHH	TTHHH	TTHHT
TTHT	TTHTH	TTHTT
TTTH	TTTHT	TTTHT
TTTT	TTTTH	TTTTT

Using the outcomes for 5 coins and a 6th coin, the sample diagram for 6 coins will have 64 sample points.

Example 5

Tossing 2 dice has 36 sample points.

The possible sums are 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12.

Table for sums and their probabilities.

X = x	2	3	4	5	6	7	8	9	10	11	12
P(X = x)	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Possible products are:

1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 16, 18, 20, 24, 25, 30, 36

Possible products and probability of getting them.

X = x	**P(X = x)**
1	$\frac{1}{36}$
2	$\frac{2}{36}$
3	$\frac{2}{36}$
4	$\frac{3}{36}$
5	$\frac{2}{36}$
6	$\frac{4}{36}$
8	$\frac{2}{36}$
9	$\frac{1}{36}$
10	$\frac{2}{36}$
12	$\frac{4}{36}$
15	$\frac{2}{36}$
16	$\frac{1}{36}$
18	$\frac{2}{36}$
20	$\frac{2}{36}$
24	$\frac{2}{36}$
25	$\frac{1}{36}$
30	$\frac{2}{36}$
36	$\frac{1}{36}$

Sample space diagram for absolute difference

1^{st} die

2^{nd} die

	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

ACTIVITY: MT/2/3-6

1. Any two events that cannot occur together are mutually exclusive
   e.g. Tom is 2 years old today and
   Tom is 6 years old today.

2. $P(A) = 1 - P(A')$
   $= 1 - 0.55$
   $= 0.45$

   Therefore $P(AUB) = P(A) = P(B) = 0.70$

3. Kato cannot be in Kibuli and Kabwangasi at the same time. Hence the two events are mutually exclusive.

4. $n(B \cap N) = 0$
   $P(B \cap N) = 0$
   B and N are mutually exclusive

5. 0
   Events are: Acan is a pupil in P.3B
   Acan is the Class Teacher of P.3B

ACTIVITY: MT/2/3-7

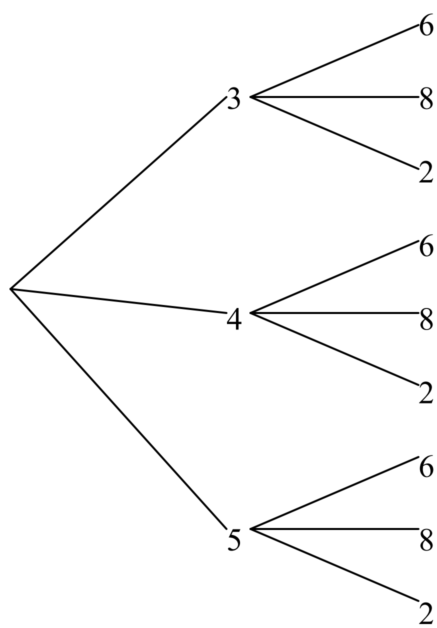
1. Any two events such that the occurrence of one does not affect the occurrence of the other.
   e.g. A: Mary's first child is a boy.
   B: Mary's second child is a girl.

   $$P(A \cap B) = \frac{1}{4}$$

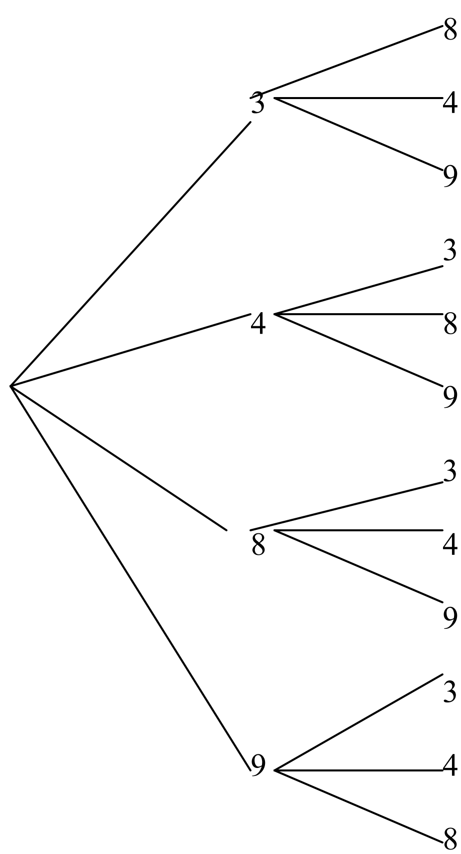
2. 0.49

3. 0.4225

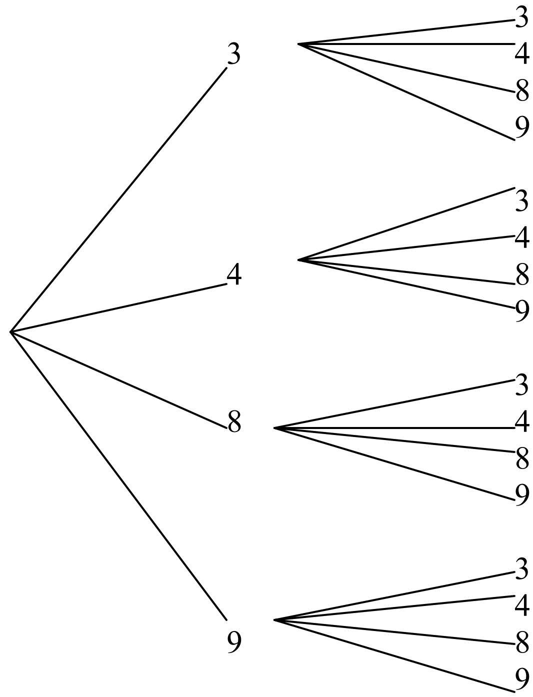
ACTIVITY: MT/2/3-8



(ii)

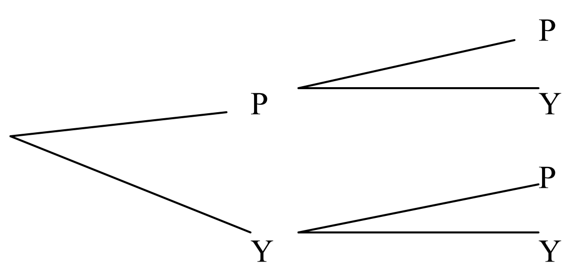


(iii)

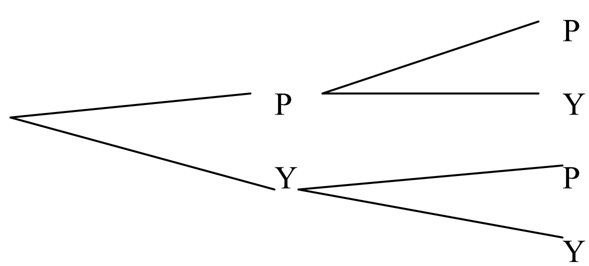


ACTIVITY: MT/2/3-9

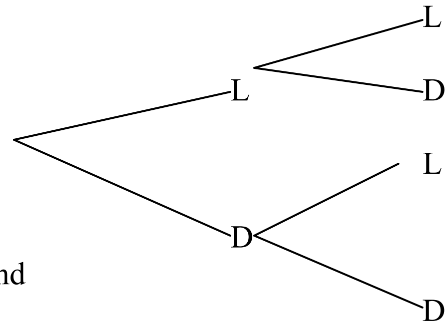
1.



2.



3.



L stands for light brown and
D stands for dark brown.

ACTIVITY: MT/2/3-10

1. (i) $\frac{35}{66}$ (ii) $\frac{31}{66}$

2. (a) $\frac{1}{84}$ (b) $\frac{1}{12}$ (c) $\frac{10}{21}$

3. (a) $\frac{3}{10}$ (b) $\frac{3}{10}$

4. (a) $\frac{9}{25}$ (b) $\frac{4}{25}$ (c) $\frac{1}{5}$

5. (a) $\frac{1}{8}$ (b) $\frac{1}{8}$ (c) $\frac{3}{4}$

END OF UNIT ASSIGNMENT MT/2-3

This assignment is meant to help you consolidate on what you have learnt in this unit. Before attempting it, please go back and revise the whole unit. Attempt all the questions before looking at the answers.

1. Two dice are tossed and the scores on their top faces are multiplied. Find the probability that the product obtained is:
   (i) equal to 24
   (ii) less than 24
   (iii) greater than 24
   (iv) equal to 25
   (v) equal to 26

   What can you say about the sum of the probabilities in (i), (ii), and (iii)?

2. A tin has 3 Fanta (F) bottle tops and 4 Mirinda (M) bottle tops. A child picks 3 bottole at random. If the child does not replace any bottle top before picking the next one, use a probability tree to find the probability of picking in the following order.

   (a) Fanta then Mirinda, then Fanta
   (b) Fanta, Fanta, Mirinda
   (c) Mirinda, Fanta, Mirinda
   (d) Only one Mirinda top
   (e) All three are Fanta tops

3. Two bags A and B each contains 5 ball point pens. A has two red and three blue pens and B has three red and two blue pens. One bag is selected at random and a pen picked from it. Find the probability that:

   (i) The pen is red.
   (ii) The pen is blue.

4. A number is chosen at random from the set.

   S = (1, 2, 3,…….., 40)

   Find the probability that the number chosen:
   (i) is divisible by 9
   (ii) is not divisible by 8
   (iii) is divisible by 3 or 5 or both.
   (iv) is divisible by 4 and also divisible by 6.
   (v) Is a multiple of 2 and a multiple of 7. What can you say about these two events?

5. Two events A and B are defined as follows:

A: Nandudu votes for Museveni in 2001.
B: Nandudu's first child is a boy.

Given that $P(A) = \frac{1}{2}$, find $P(A \cap B)$. What can you say about the two events A and B?

6. Box A contains 8 tomatoes and another box B contains 6 oranges. One fruit is picked from each box. Draw a sample space diagram for the possible outcomes, given that each tomato and each orange is different from all the others.

7. C is the event that the Headmaster is in his office and D is the event that the Headmaster has gone to see the LC V Chairperson. Given that P© = 0.47 and $P(D) = 0.28$, use a venn diagram to find:

(i) $P(C' \cap D')$ (II) $P(C \cup D)$
What does each of the probabilities in (i) and (ii) represent?

8. When electing the Youth MP, A is the event of electing a University graduate and B is the event of electing a member of the Young Movementists.

If $P(A') = 0.41$, $P(AUB) = 0.66$ and $P(B) = 0.22$, use a venn diagram to find:

$P(A \cap B)$, $P(A \cap B')$, $P(A' \cap B)$, $P(A'UB)$, $P(A' \cap B')$ and $P(A' \cup B')$.

What does each of the events above represent in terms of the MP elected?

LEARNING OUTCOMES SELF CHECKING EXERCISE

You have now come to the end of unit 3 of Module MT/2. The learning outcomes are listed below. Put a tick in the column which best reflects your learning.

	Learning Outcomes	***Satisfactory***	***Not Sure***
1.	I can solve probability problems using set theory.		
2.	I can solve probability problems using probability trees.		
3.	I am able to identify mutually exclusive and independent events.		
4.	I can solve problems involving mutually exclusive and independent events.		

If you placed a tick in the **"NOT SURE"** column, please go back to the text and reinforce your learning. Remember it is worth the effort.

ANSWERS TO END OF UNIT ASSIGNMENT MT/2-3

1. Sample space is:

1st die

2nd die

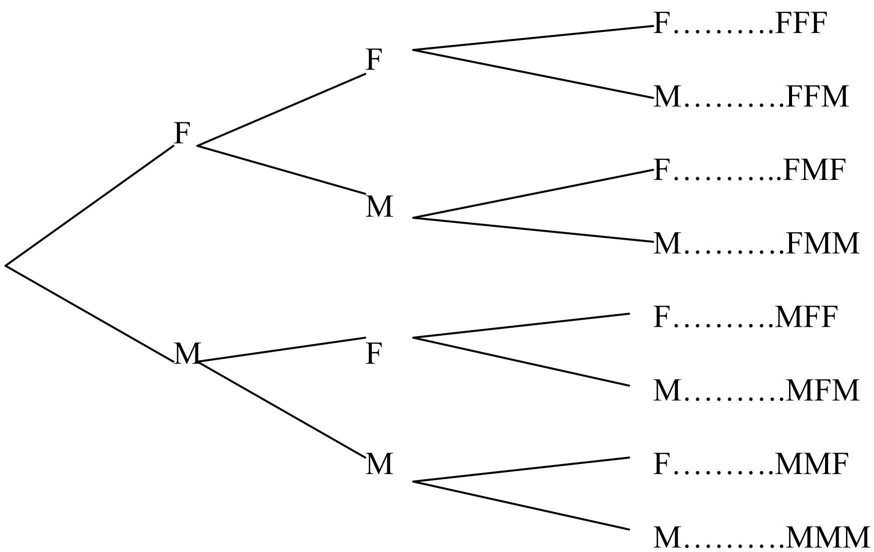
	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

(i) $\frac{1}{18}$ (ii) $\frac{5}{6}$ (iii) $\frac{1}{9}$ (iv) $\frac{1}{36}$ (v) 0

Probabilities $\frac{1}{8} + \frac{5}{6} + \frac{1}{9} = \frac{1+15+2}{18} = 1$

Because an outcome is either equal to 24, less than 24 or greater than 24. These three events make up eh sample space.

2. The probability tree is as below:



(a) P(F, M, F) $= \frac{3}{7} \times \frac{4}{6} \times \frac{2}{5} = \frac{4}{35}$

(b) P(F, F, M) $= \frac{3}{7} \times \frac{2}{6} \times \frac{4}{5} = \frac{4}{35}$

(c) P(M, F, M) $= \frac{4}{7} x \frac{3}{6} \text{ x } \frac{3}{5} = \frac{6}{35}$

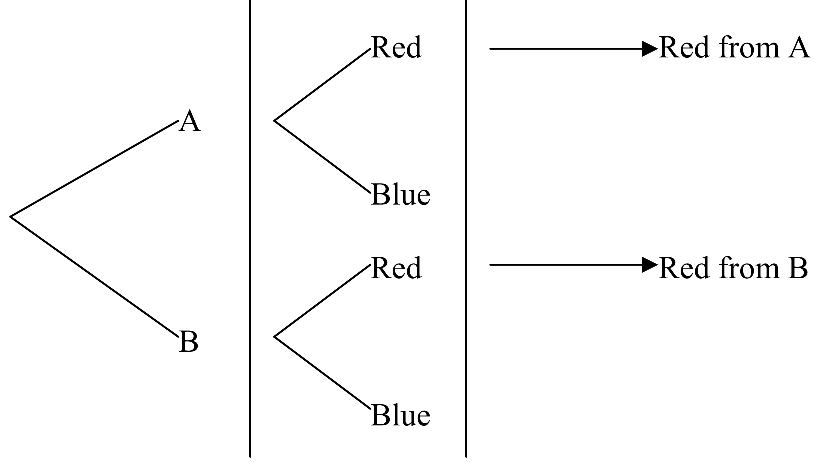
(d) P(only one Mirinda top)

$= \text{P(FFM)} + \text{P(FMF)} + \text{P(MFF)}$

$= \frac{4}{35} + \frac{4}{35} + \frac{4}{35} = \frac{12}{35}$

(e) P(All three Fanta tops) = P(FFF) $= \frac{3}{7} \text{ x } \frac{2}{6} \text{ x } \frac{1}{5} = \frac{1}{35}$

3. Choice of bag | Choice of pen



(i) P(Red) = P(Red from A or Red from B)
= P(Red from A + P(Red from B)

$= \frac{1}{2} \text{ x } {}^{2}\!/_{5} + \frac{1}{2} \text{ x } {}^{3}\!/_{5}$

$= \frac{2+3}{10}$

$= \frac{5}{10}$

$= \frac{1}{2}$

(ii) P(Blue) $= \frac{1}{2} \text{ x } \frac{3}{5} + \frac{1}{2} \text{ x } \frac{2}{5} = \frac{1}{2}$

4. S = $\{1,2,3,4,....40\}$

(i) Let D be the event that number chosen is divisible by g.

Then D = {9, 18, 27, 36}

$$P(D) = n(D) = \frac{4}{40} = \frac{1}{10}$$

(ii) Let N be the vent that number chosen is divisible by 8
Then N = {8, 16, 24, 32, 40}. nN) = 5;
Therefore $n(N') = \frac{35}{40} = \frac{7}{8}$

(iii) C is the event number that is divisible by 3 or 5 or both,
Then C = {3, 5, 6, 9, 10, 12, 15, 18, 20, 21, 24, 25, 27, 30, 33, 35, 36, 39}

nC) = 18

$$\text{Therefore } P(C) = \frac{18}{40} = \frac{9}{20}$$

(iv) F is even number that is divisible by 4 and also divisible by 6;
F = {12, 24, 36}

$$P(F) = \frac{3}{40}$$

(v) 0. the events are mutually exclusive.

5. $$P(A\cap B) = \frac{1}{2} \text{ x } \frac{1}{2} = \frac{1}{4}$$

The events are independent.

6. Let the tomatoes be labeled.
IT, 2T, 3T, 4T, 5T, 6T, 7T, 8T and the oranges be IG, 2G, 3G, 4G, 5G, 6G.

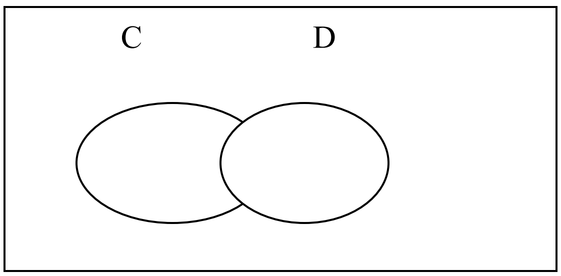
The samples space is:

Tomatoes

Oranges

	IT	2T	3T	4T	5T	6T	7T	8T
1G	1GIT	1G2T	1G3T	1G5T	1G5T	1G6T	1G7T	1G8T
2G	2GIT	2G2T	2G3T	2G4T	2G5T	2G6T	2G7T	2G8T
3G	3GIT	3G2T	3G3T	3G4T	3G5T	3G6T	3G7T	3G8T
4G	4GIT	4G2T	4G3T	4G4T	4G5T	4G6T	4G7T	4G8T
5G	5GIT	5G2T	5G3T	5G4T	5G5T	5G6T	5G7T	5G8T
6G	6GIT	6G2T	6G3T	4T6G	6G5T	6G6T	6G7T	6G8T

7.



Clearly, C and D are mutually exclusive.

Therefore $P(C \cap D) = 0$

$P(S) = 1$

Therefore $x + 0.47 + 0.28 = 1$

$x = 1 - 0.47 - 0.28$

$= 0.25$

(i) $P(C' \cap D') = 0.25$

This is the probability that the Headmaster is not in his office and he has not gone to see the LC V Chairperson.

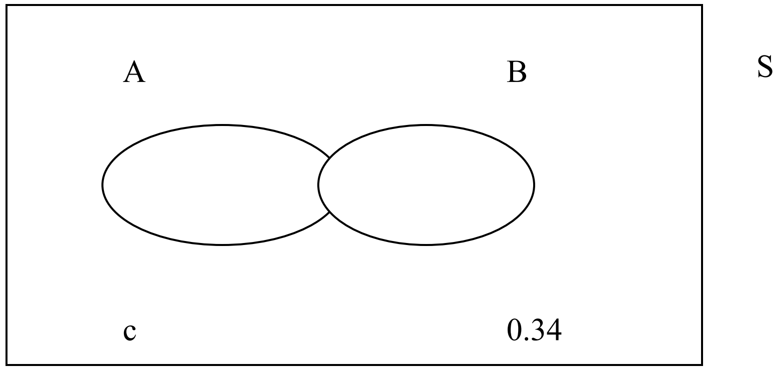
(ii) $P(C \cup D) = P(C) + P(D)$

$= 0.47 + 0.28$

$= 0.75$

This is the probability that the Headmaster has gone to see the LC V Chairperson or he is in his office.

8.



$$P(A) = 1 - P(A') = 1 - 0.41 = 0.59$$

Then $a = 0.59 - x$ ……………………………………… (i)

$b = 0.22 - x$ ………………………………………..(ii)

$c = 1 - (a + b + x)$ …………………………………(iii)

But from $P(AUB) = 0.66$

$a + b + x = 0.66$ ……………………………………..(iv)

Substituting for a and b, from (i) and (ii), (iv) becomes

$$(0.59 - x) + (0.22 - x) + x = 0.66$$
$$0.81 - x = 0.66$$
$$x = 0.15$$

Then $a = 0.44$

$b = 0.07$

$c = 0.34$

$P(A\cap B) = 0.15$; the probability that the MP elected is a University graduate and a Young Movement.

$P(A\cap B') = 0.44$; MP elected is a University graduate but not Young Movement

$P(A'\cap B) = 0.07$; MP elected is not a University graduate but is Young Movement.

$P(A'UB) = 0.41$; MP elected is not a University graduate or is a Young Movementist or both.

$P(A'UB')$ = 0.34; MP is neither a University graduate nor a Young Movementist.

$P(A'UB')$ = 0.85; MP is not a University graduate or is not a Young Movementist or both.

REFERENCES

1. School Mathematics of East African Book 3, Cambridge University Press.

2. School Mathematics of East African, Books 4.

3. Davis B.K. Kamoga, Ordinary Level Model Mathematics with questions, Solutions and Notes, 1999.

4. N.M. Patel Mathematics for Kenya Schools Form 2 and 4. Malim Publications.

5. Murray R. Spiegel, Schanm's Outline Series, Theory and Problems of Statistics.

6. Greer A. (1980), A First Course in Statistics. England, Stanley Thornes (Publishers) Ltd.

7. Rees D.G. (1987), Foundations of Statistics. London, Chapman and Hall.

8. VUP (1971), School Mathematics of East African, Book 3, Press Syndicate, Melbourne.

9. Secondary School Mathematics Students' Book 3, Macmillan Uganda, Kampala.