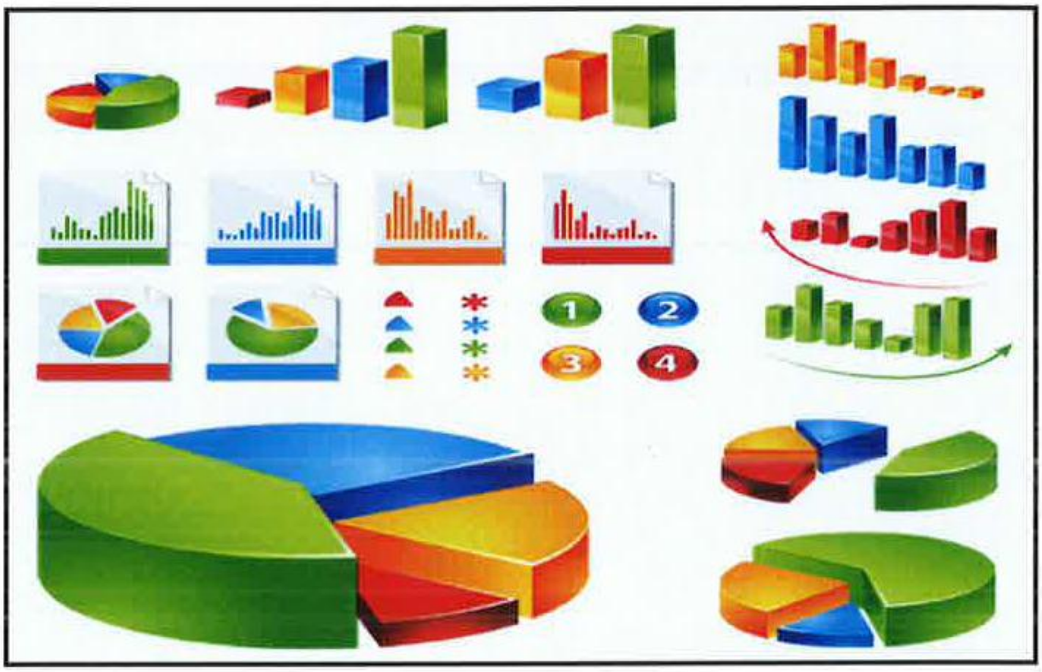
DEPARTMENT OF EDUCATION

GRADE 9

MATHEMATICS

UNIT 3



WORKING WITH DATA

Published by:

FLEXIBLE OPEN AND DISTANCE EDUCATION
PRIVATE MAIL BAG, P.O. WAIGANI, NCD
FOR DEPARTMENT OF EDUCATION
PAPUA NEW GUINEA

MATHEMATICS

GRADE 9

UNIT 3

WORKING WITH DATA

TOPIC 1:	**ORGANIZATION OF DATA**
TOPIC 2:	**PRESENTATION OF DATA ON GRAPHS**
TOPIC 3:	**MEASURES OF CENTRAL TENDENCY**
TOPIC 4:	**MEASURES OF SPREAD**

Acknowledgements

We acknowledge the contribution of all Secondary and Upper Primary teachers who in one way or another helped to develop this Course.

Special thanks are given to the staff of the Mathematics Department- FODE who played active role in coordinating writing workshops, outsourcing of lesson writing and editing processes involving selected teachers in Central and NCD.

We also acknowledge the professional guidance and services provided through-out the processes of writing by the members of:

Mathematics Department- CDAD
Mathematics Subject Review Committee-FODE
Academic Advisory Committee-FODE

.
This book was developed with the invaluable support and co-funding of the GO-PNG/FODE World Bank Project.

Flexible Open and Distance Education
Papua New Guinea

Published in 2016

ISBN: 978-9980-87-732-1

National Library Services of Papua New Guinea

Written and compiled by: Luzviminda B. Fernandez
Senior Curriculum Officer
Mathematics Department
FODE

Printed by the Flexible, Open and Distance Education

CONTENTS

SECRETARY'S MESSAGE

Achieving a better future by individuals students, their families, communities or the nation as a whole, depends on the curriculum and the way it is delivered.

This course is part and parcel of the new reformed curriculum – the Outcome Base Education (OBE). Its learning outcomes are student centred and written in terms that allow them to be demonstrated, assessed and measured.

It maintains the rationale, goals, aims and principles of the national OBE curriculum and identifies the knowledge, skills, attitudes and values that students should achieve.

This is a provision of Flexible, Open and Distance Education as an alternative pathway of formal education.

The Course promotes Papua New Guinea values and beliefs which are found in our constitution, Government policies and reports. It is developed in line with the National Education Plan (2005 – 2014) and addresses an increase in the number of school leavers which has been coupled with limited access to secondary and higher educational institutions.

Flexible, Open and Distance Education is guided by the Department of Education's Mission which is fivefold;

- to facilitate and promote integral development of every individual
- to develop and encourage an education system which satisfies the requirements of Papua New Guinea and its people
- to establish, preserve, and improve standards of education throughout Papua New Guinea
- to make the benefits of such education available as widely as possible to all of the people
- to make education accessible to the physically, mentally and socially handicapped as well as to those who are educationally disadvantaged

The College is enhanced to provide alternative and comparable path ways for students and adults to complete their education, through one system, many path ways and same learning outcomes.

It is our vision that Papua New Guineans harness all appropriate and affordable technologies to pursue this program.

I commend all those teachers, curriculum writers and instructional designers, who have contributed so much in developing this course.

....

DR. UKE KOMBRA, PhD
Acting Secretary for Education

UNIT 3: WORKING WITH DATA

Dear Student,

This is Unit 3 of the Grade 9 Mathematics Course. It is based on the NDOE Lower Secondary Mathematics Syllabus and Curriculum Framework for Grade 9 as part of the continuum of Mathematics knowledge and learning from Grade 7 to 10.

This Unit consists of four Topics:

Topic 1: Organization of Data
Topic 2: Presentation of Data
Topic 3: Measures of Central Tendency
Topic 4: Measures of Spread

In Topic 1- **Organization of Data**-You will identify the different types of data and learn to organize the different types of data using frequency distribution tables, stem and leaf plots.

In Topic 2- **Presentation of Data**- You will learn further about the different graphs and charts to help you illustrate different types of data such as pictograph, bar graphs, column graphs, histograms and frequency polygons. You will also learn about cumulative and relative frequencies and their graphs.

In Topic 3- **Measures of Central Tendency**- You will look at the mean, median and mode of grouped and un-grouped sets of data and learn how to calculate them. You will also learn the conditions under which it is most appropriate to use each of them. .

In Topic 4- **Measures of Spread**- You will learn to find the range of ungrouped and grouped sets of data

The Topics are divided into 5 to 6 lessons. Each lesson provides you with reading materials showing worked examples and practice exercises. The answers to the practice exercises are given at the end of each topic.

A study guide is also provided to assist you in studying this unit.

We hope that you will find this strand both challenging and interesting.

All the best!

Mathematics Department
FODE

STUDY GUIDE

Follow the steps given below as you work through the Unit.

Step 1: Start with TOPIC 1 Lesson 1 and work through it.

Step 2: When you complete Lesson 1, do Practice Exercise 1.

Step 3: After you have completed Practice Exercise 1, check your work. The answers are given at the end of TOPIC 1.

Step 4: Then, revise Lesson 1 and correct your mistakes, if any.

Step 5: When you have completed all these steps, tick the check-box for the Lesson, on the Contents Page (page 3) like this:

$\boxed{\surd}$ Lesson 1: Types of data

Then go on to the next Lesson. Repeat the process until you complete all of the lessons in Topic 1.

> As you complete each lesson, tick the check-box for that lesson, on the Content Page 3, like this $\boxed{\surd}$.This helps you to check on your progress.

Step 6: Revise the Topic using Topic 1 Summary, then, do Topic test 1 in Assignment 2.

Then go on to the next Topic. Repeat the same process until you complete all of the four Topics in Unit 2.

Assignment: (Four Topics and a Unit Test)
When you have revised each Topic using the Topic Summary, do the Topic Test in your Assignment. The Unit book tells you when to do each Topic Test.

When you have completed the four Sub-strand Tests, revise well and do the Strand test. The Assignment tells you when to do the Strand Test.

> The Topic Tests and the Unit test in the Assignment will be marked by your Distance Teacher. The marks you score in each Assignment will count towards your final mark. If you score less than 50%, you will repeat that Assignment.

Remember, if you score less than 50% in three Assignments, you will not be allowed to continue. So, work carefully and make sure that you pass all of the Assignments.

TOPIC 1

ORGANIZATION OF DATA

TOPIC 1: ORGANIZATION OF TYPES OF DATA

Introduction

Statistics is the name given to the science of collecting, organizing, presenting and analysing data. After data are collected, they are arranged and organized so that they can be easily understood.

Once the data or information has been chosen and the data are collected, it is important that they are summarized and presented in a method in which it is easy to understand and visualize.

For example the table below is the frequency table displaying the data or information about the height of Grade 9 Students..

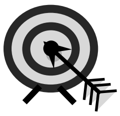
Height (cm)	Tally Marks	Frequency
155	/	1
156	///	3
157	~~////~~	5
158	////	4
159	~~////~~ - ////	9
160	~~////~~ - /	6
161	~~////~~ - ///	8
162	~~////~~ - //	7
163	//	2
	Total:	45

In this topic, you will:

- Identify the different types of data
- define and identify the features of a frequency distribution
- organize raw data in a frequency distribution table
- describe stem and leaf plot and identify the steps in making them and use them to organize and display data.
- construct frequency distribution table for discrete and continuous data
- define and draw grouped frequency distribution table.

Lesson 1: Types of Data

You have learned something about data in your earlier study of Grade 7 and 8 Mathematics.



In this lesson, you will:

- revise the meaning of data
- identifiy the types of data

Arranging information so that it can be easily understood is called **organizing data.**

Data is another name for information or group of facts.

Vast amount of raw data are being collected all the time.

Raw data is information that has not been ordered or processed in any way.

Data can be classified as:

1. Qualitative or Categorical (non- numerical data)
2. Quantitative (numerical data)

Qualitative or Categorical data describes characteristics or qualities that cannot be counted.

For example: The texture, colour, gender are properties that are not numbers.

Quantitative data describes characteristics that has numerical value and can be counted or measured.

For example: the number of books in a shelf, the height of a person, the weight of a student.

Further, Quantitative data can be either **discrete** or **continuous.**

Discrete data are data that take exact numerical values. It is often the result of counting. It is usually concerned with a limited number of countable values and cannot take the form of decimals.

An example is the size of a particular family since it can only take a specific value such as 1,2,3,4 and so on. Values between them like 1.5 or 3.2 are not possible. We cannot have a family with 5.5 members.

Here are other examples of discrete data.

1. shoe size
2. marks in a test
3. number of students in a class
4. number of goals scored by a netball team
5. number of cars sold per week by a car company

Continuous data are measured on some scale and can take any value within that scale. It is usually the result of measuring.

For example, if the weight of the student is given as 48 kg, the exact weight could be anywhere between 47.5 and 48.5 kg. Weight is a continuous data.

Here are other examples of continuous data.

1. Height
2. Length
3. Width,
4. Time
5. Amount of rainfall in each month per year
6. Amount of sunshine in a day

When collecting data, we are interested in a particular property or characteristic of a group of people or objects. This particular characteristic that we are interested in is called a **variable.**

A variable is a property able to assume different values.

For example, temperature is a variable. Data can be collected on it.

Now look at the following examples of classifying data.

Example 1

Classify the following data as categorical, discrete or continuous.

1. The number of heads when 3 coins are tossed.
2. The brand of toothpaste used by students in a class
3. The heights of a group of 16 years old children

Answers:

1. The values of the data are obtained by counting the number of heads. The result can only be one of the exact values 0, 1, 2, or 3.
   It is a **discrete data**.

2. The variable describes a brand of toothpaste. It is **categorical data**.
3. It is a numerical data obtained by measuring. The results can take any value between certain limits determined by the degree of accuracy of the measuring device. It is **continuous data.**

Example 2

Sam buys a new dress.

Write down two variables associated with a dress that shows the following data types:

(a) Qualitative

(b) Discrete

(c) Continuous.

Answers:

(a) Colour and texture of the material are **qualitative**

(b) The size of the dress and the number of buttons it has are **discrete**

(c) The length of the sleeves and the diameter of each button are **continuous.**

Example 3

Are the following variables discrete or continuous?

(a) Volume of a bottle,

(b) Number of radios produced in a day,

(c) Number of people absent from work on a workday

(d) Average number of pawpaw harvested.

Answer.

(a) As volume can take decimal values, it is **continuous**

(b) As this is a count, it will be a whole number, it is **discrete.**

(c) As this is also a count, it will be **discrete**

(d) This is not a count but an average of counts, so this can take decimal values. It is therefore **continuous**..

NOW DO PRACTICE EXERCISE 1

Practice Exercise 1

1. For each of the following investigations, classify the variable as categorical, discrete or continuous

   (a) the number of people who die from HIV/AIDS each year
   (b) the heights of the members of a rugby team
   (c) the most popular sports
   (d) the number of children in a New Guinean family
   (e) the fuel consumption of different cars
   (f) the marks scored in a mathematics tests
   (g) the pulse rates of a group of athletes
   (h) the most popular colour of cars
   (i) the gender of school principals
   (j) the time spent doing assignments
   (k) the amount of rainfall in each months of the year
   (l) the items sold in a school canteen
   (m) the reasons people pay taxes
   (n) the number of matches in a box
   (o) the pets owned by a class of students

---

2. Kila is spending the holiday hiring out deck chairs at the beach.

   (a) Is the number of deck chairs hired out each day a discrete or continuous variable?

   (b) Describe a qualitative variable associated with the deck chairs.

---

3. Sort the following into (i) discrete (ii) continuous and (iii) categorical data

   (a) The weight of a parcel and the cost of its postage

   (b) The number of cups of sugar and the amount of sugar needed in a cake recipe

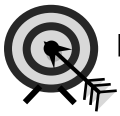
   (c) How long will you take to finish in a cross country race and your finishing position in the race

---

CORRECT YOUR WORK. ANSWERS ARE AT THE END OFTOPIC 3

Lesson 2: Frequency Distribution of Categorical Data

You've learnt the meaning of data and identified the different types of data in the previous lesson.



In this lesson, you will:

- define and identify the features of a frequency distribution
- organize raw data on a frequency distribution table.

Once a sample has been chosen and data are collected, it is necessary to find some means of organizing them and describing the data obtained from the study.

Data are often collected in an unorganized and random manner. Before we can draw conclusions from them, they must be summarized and represented in a way that is easy to visualize and understand.

Arranging information so that it can be easily understood is called **organizing data.**

We can organize the data in a frequency table.

Frequency is the term used for the number of times a particular score occurs in a set of data.
A **frequency table** is a table used to set out numerical information, so that the information is easily read and understood.

The arrangement of data showing the frequency with which a measure of a given size occurs is called **frequency distribution.**

Earlier in Lesson 1, you learnt the meaning of categorical data. As you have learnt, categorical data are data which describes a characteristic or quality that cannot be counted. It can be divided into categories.

When we tabulate the categorical data into a frequency distribution table, the table is headed by a **number** and a **title** to give the reader an idea of the nature of the data being organized. For example, “Men and Women Majoring in Mathematics” is the title and the number you can assign to the table may be 1 or 2.1 as the case may be.

For this type of data, our frequency table should consist of two columns as presented in Table 2.1. See next page.

The first column pertains to the characteristic being presented and contains the categories of analysis. In the given example, sex is the characteristic being presented, whose levels are called the **categories of analysis**.

The second column is headed by “f”, the frequency consisting of the number of subjects in each category as well as the sum of all the number of subjects which is 130.

Table 2.1
MEN AND WOMEN MAJORING IN MATHEMATICS AT UPNG

Sex	**Frequency (f)**
Men Women	23 107
Total	130

Now look at the example below on how to make a frequency distribution table of categorical data.

Example 1

The method by which the employees of a certain company travelled to office on a particular day is recorded below, using the following codes: Walk (W), Taxi (T), Bus (B), Private Car (P), and Company Car(C).

WTBPT BBBWB BBTBP TCTBP PPBPP PCCTB

Rearrange this information into a frequency distribution table using tally column.

Solution:

Table 2.2
METHOD BY WHICH COMPANY EMPLOYEES TRAVELLED TO OFFICE

Method of Travel	**Tally Marks**	**Frequency (f)**
Walk (W)	*II*	2
Taxi (T)	*~~IIII~~ – I*	6
Bus (B)	*~~IIII~~ – ~~IIII~~ – 1*	11
Private Car (P)	*~~IIII~~ – III*	8
Company Car (C)	*III*	3
	Total	30

Steps:

(1) List all the codes (methods of travel) in the first column. From the above list we have: Walk (W), Taxi (T), Bus (B), Private Car (P), Company car (C).

(2) Read through the list of codes. Each time a code occurs put a tally mark, which is a stroke (***I***) against the code. To make counting code easier the tally marks are grouped in fives (***~~IIII~~***), the fifth stroke being drawn diagonally across the first four.

(3) When we have been through the list of codes, we count the tally marks for each code. This gives the frequency for each code. The frequency is the total of the tally marks, that is, the number of times a particular mode of travel is used. (see above)

(4) Always check that the total frequency column is the same as the number of observations recorded.

Example 2

The colours of cars passing the front of a school in a 30 minute period are recorded below using the codes: white (W), blue (B), grey (G), red (R), others (O)

BRGWO	BWROW	BGRWW	GBRWO	GBRWG
BRGOW	BWGRB	WWBRG	WBRWB	BRRGW

(a) Rearrange this data into a frequency distribution table using tally marks.
(b) How many cars passed the front of the school in this time period?
(c) What was the most popular car colour in this survey?

Solution:

(a) **Steps:**

(1) List all the codes (car colours) in the first column. From the above list we have: White (W), blue (B), grey (G), red (R) and others (O).

(2) Read through the list of codes. Each time a code occurs put a tally mark, which is a stroke (***/***) against the code. To make counting code easier the tally marks are grouped in fives (***IIII̸***), the fifth stroke being drawn diagonally across the first four.

(3) When we have been through the list of codes, we count the tally marks for each code. This gives the frequency for each code. The frequency is the total of the tally marks, that is, the number of times a car with a particular colour passes by.

(4) Always check that the total frequency column is the same as the number of observations recorded.

Table 2.3
COLOUR OF CARS PASSING
THE FRONT OF A SCHOOL IN 30 MINUTES

Colour of Cars	**Tally Marks**	**Frequency (f)**
white (W)	*IIII̸ – IIII̸ - IIII*	14
blue (B)	*IIII̸ – IIII̸ – II*	12
green (G)	*IIII̸ – IIII*	9
red (R)	*IIII̸ – IIII̸ - I*	11
others (O)	*IIII*	4
	Total	50

(b) 50 cars

(c) White car

NOW DO PRACTICE EXERCISE 2

Practice Exercise 2

1. Sam was tasked to find out how many of his classmates chose English, Science, Mathematics, Social Science, Personal Development and Design and Technology as their favourite subjects. His result was recorded as shown.

E E E E E E E E E E E
S S S S S S S S S S
M M M M M M M M M M
SS SS SS SS SS
PD PD PD PD PD PD PD
DT DT DT DT DT DT DT DT

(a) Rearrange this data into a frequency distribution table using tally marks.

(b) What title will you give the table?

(c) How many students like Personal Development (PD)?

(d) What subject is the most favourite?

2. A survey was done to find the brand of a car owned by a group of people. The results of the survey are recorded below using the code:

Ford (F), Mazda (M), Suzuki (S), Toyota (T), Honda (H), Nissan (N)

FMSTTH	MMSSTT	MMMTTT	FMMSSTH	MSSSTT
MSSSST	MTTTHH	TTTHHN	TTTTTHS	TTHHNN

a) Rearrange this data into a frequency distribution table using tally marks.

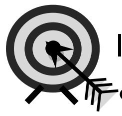
b) How many people were surveyed?

c) What was the most popular car in this survey?

CORRECT YOUR WORK. ANSWERS ARE AT THE END OFTOPIC 1

Lesson 3: Frequency Distribution of Discrete Numerical Data

You've defined frequency distribution and learnt to organize raw data on a frequency distribution table.



In this lesson, you will:

- revise discrete data and frequency tables
- organize discrete data in a frequency distribution table.

As you have learnt in the previous lesson, discrete data are data which can only take whole number or exact numerical values. When we count things, the answers we get are whole numbers.

These are the most common examples of discrete data.

(a) Number of people who use a micro- computer in an hour

(b) The number of cars sold in a day

(c) Number of radios produced in a day

(d) Number of students absent in a class

(e) Number of children in a family

(f) Number of mistakes in a test and so on.

When we have a set of raw data we usually wish to summarize the figures into something more manageable and easily to understand. Our first step is often to put the data values into their numerical order.

For example a group of 50 students was given a spelling test and a number of mistakes for each student were recorded as follows:

1	5	0	2	4	5	2	3	3	0
3	2	3	1	3	3	2	3	2	0
3	3	3	1	2	2	1	2	2	4
0	1	3	3	3	2	2	4	1	1
5	4	3	2	3	3	3	3	1	0

This information can be presented in a frequency distribution table, or more simply a frequency table. To draw a frequency table

(i) List all possible scores in one column, the first row of the column having the lowest score, the last having the highest.
For the list above, these scores are 0, 1, 2, 3, 4, 5.

(ii) Read through the list of scores. Each time a score occurs put a tally mark, which is a stroke (/) against the score. To make counting the scores easier the tally marks are grouped in fives (~~////~~), the fifth stroke being drawn across the first four.

(iii) When we have read through the list of scores, we count the tally marks for each score. This gives the frequency for each score. (See table below)

(iv) Construct the frequency table to display the data.

Number of Mistakes (Scores)	**Tally Marks**	**Frequency**
0	////	5
1	~~////~~ - ////	8
2	~~////~~ - ~~////~~ - //	12
3	~~////~~ - ~~////~~ - ~~////~~ - ///	18
4	////	4
5	///	3
	Total:	50

This is a frequency table of individual scores. The total frequencies should always be checked to make sure it is the same as the number of original scores.

Example 2

Stephen asked students in his class to indicate how many pets they had. This resulted in the following data.

1	3	2	2	4	1	5	2	1	1
6	4	1	2	5	2	1	4	1	2

For this data, draw the frequency distribution table that shows the number of pets the students had.

Solution:

(i) List all possible scores in one column, the first row of the column having the lowest number of pets, the last having the highest.
For the list above, these numbers of pets are 1, 2, 3, 4, 5. 6

(ii) Read through the list of numbers. Each time a score occurs put a tally mark, which is a stroke (/) against the score. To make counting the scores easier the tally marks are grouped in fives (~~////~~), the fifth stroke being drawn across the first four.

(iii) When we have read through the list of scores, we count the tally marks for each score. This gives the frequency for each score.

(iv) Construct the frequency table to display the data.

Number of Pets (Scores)	**Tally Marks**	**Frequency**
1	~~////~~- //	7
2	~~////~~ - /	6
3	/	1
4	///	3
5	//	2
6	/	1
	Total:	20

The table shows the frequency of each number of pets. The total frequencies should always be checked to make sure it is the same as the number of original data..

Example 3

For a class of 25 students the following marks out of 10 were obtained in a test.

5	4	6	6	5	3	9
9	8	10	3	6	7	3
4	5	6	5	7	10	7
6	7	8	9	4		

If this information is organized in a frequency distribution table, it looks like this:

Marks (Scores)	**Tally Marks**	**Frequency**
3	///	3
4	///	3
5	////	4
6	~~////~~	5
7	////	4
8	//	2
9	///	3
10	//	2
	Total	26

Remember a frequency distribution table is very good for collecting and organizing data, but when analysing data it is often more desirable to have the information presented in the form of diagram or graph.

NOW DO PRACTICE EXERCISE 3

Practice Exercise 3

1) The trees in each backyard of Waigani Village Houses were counted and the number recorded. The data is shown below.

7	6	12	2	0	4	6	3	3	5
8	5	9	1	4	6	4	8	1	7
2	5	3	4	2	1	3	4	5	1
3	5	2	2	0	3	3	2	7	1
5	10	5	4	4	2	6	1	4	5

(a) What are the highest and lowest scores in this data?

(b) Organize this data in a frequency distribution table.

2. A goal kicker for a football team kicked the following number of goals in his twenty-four games in the last season.

2	2	1	1	4	2	3	0
3	1	0	6	4	1	2	3
2	0	2	5	1	5	4	1

Complete a frequency distribution table for this set of data.

3. Two dice were thrown one hundred times and the total showing on the two upper dice was recorded to obtain this set of score.

4	6	9	6	5	11	7	5	9	8
5	3	4	7	9	10	12	8	10	4
9	6	7	5	10	8	9	11	3	7
7	5	8	10	11	7	10	9	11	6
12	3	9	4	5	7	3	5	6	2
2	8	8	7	9	6	8	4	8	8
10	5	6	8	2	10	5	6	7	4
6	4	7	8	6	7	9	7	9	7
5	7	5	8	9	6	8	7	10	6
7	6	8	4	5	7	3	8	6	4

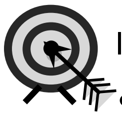
(a) What are the highest and lowest scores in this data?

(b) Organize this data in a frequency distribution table.

CORRECT YOUR WORK ANSWEWRS ARE AT THE END OF TOPIC 1

Lesson 4: Stem and Leaf Plots

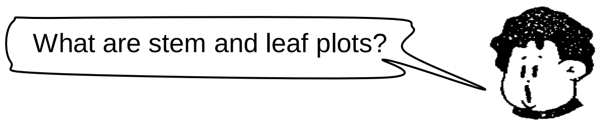
You have revised discrete data and frequency distribution table in the previous lesson.



In this lesson, you will:

- define stem and leaf plots
- identify steps in making a step and leaf plot
- use stem and leaf plot to organize and display data.

Another way of displaying information is the Stem and Leaf Plots. It is used to group and rank data to show the range and distribution of the data.



Stem and leaf plot or stem plot is a diagram that shows all the original data and also gives the original picture or trend for the data.

You can use stem and leaf plots to display discrete and continuous data.

In a stem and leaf plot, the values are grouped so that all but the last digit is the same in each category. For two-digit numbers, the tens values are the stem and the units are the leaves.

Example 1

Given below are the results obtained by 23 students in a Mathematics test.

54	75	63	80	63	77	78	86
72	62	94	84	87	66	93	56
80	86	51	78	68	73	82	

Show this data using a stem and leaf plot.

Solution:

Stem	Leaf
5	1 4 6
6	2 3 3 6 8
7	2 3 5 7 8 8
8	0 0 2 4 6 6 7
9	3 4

This row represents the numbers 51, 54 and 56.

Scores ranges from 51 to 94, so stems are 5 to 9.

Key: 5 | 1 means 51

Example 2

The results in an English class test out of 70 are given below.

55	43	46	66	45	57
22	42	65	41	65	63
23	70	53	57	45	65
26	48	46	23	61	67
51	62	57	70	55	46

(a) Draw a stem and leaf plot to represent this data.

(b) What are the lowest and highest score?

(c) How many students scored 46?

(d) How many students scored a mark in the sixties?

Solution:

(a) In this stem and leaf plot, the tens digit forms the stem and the units digit forms the leaf. This means that for the mark 45, the stem is the 4 and the leaf is the 5.

Stem	**Leaf**	
2	2 3 3 6	← This row represents the numbers 22, 23, 23 and 26.
3		There were no scores in the thirties.
4	1 2 3 5 5 6 6 6 8	
5	1 3 5 5 7 7 7	
6	1 2 3 5 5 5 6 7	
7	0 0	**Key:** 4 | 5 means 45

(b) Lowest Score = 22, Highest score = 70

(c) Number of students who scored 46 = 3

(d) Number of students who scored a mark in the sixties = 8

Example 3

Below is a stem and leaf plot.

0	2 5	
1	3 3 7 8	
2	0 2 6	
3	1 7	**Key**: 3 | 1 means 31

List the data values in the stem and leaf plot.

Solution: The data values are 2, 5, 13, 13, 17, 18, 20, 22, 26, 31, and 37.

Example 4

Copy and complete this table showing scores, stems and leaves.

Score	**Stem**	**Leaf**
28		
153		
91		
8		
	1	9
	2	8
	18	6
	204	9
	0	6

Solution:

Score	**Stem**	**Leaf**
28	2	8
153	15	3
91	9	1
8	0	8
19	1	9
28	2	8
186	18	6
2049	204	9
6	0	6

Note: A leaf has only one digit but a stem may have more than one digit.

Remember: With a stem and leaf plot

- all of the data is used and displayed
- the largest and smallest measurements can be found
- the clustering (grouping) of data can be more easily seen
- the length of the leaf column indicates the number of scores belonging to that stem.

NOW DO PRACTICE EXERCISE 4

Practice Exercise 4

1. The first three scores have been placed in the stem-and-leaf plot. Copy the table and add the remaining 17 scores.

Stem	Leaf
3	
4	
5	
6	

34	49	41	57	38
59	33	31	61	68
55	39	51	53	63
61	58	33	49	60

2. Draw a stem-and-leaf plot using stems of 3, 4, 5, and 6 for these 20 scores.

40	66	62	59	44	37	68	52	39	45
41	62	49	58	35	47	48	59	32	52

Stem	Leaf
3	
4	
5	
6	

3. The following stem-and-leaf plot shows the time spent (hours) watching TV by a group of students during one week.

Stem	Leaf
0	3 5 6 8 9
1	0 2 2 3 5 5 5 9
2	2 4 5 5 5 7 8
3	0 1 1 4 6

(a) How many students were surveyed?

(b) What was the least and greatest number of hours of TV a week?

(c) How many students watched less than 10 hours of TV a week?

(d) How many students watched more than 30 hours of TV a week?

4. Copy and complete this table.

Score	Stem	Leaf
39	**3**	**9**
27	**2**	**7**
125		
	8	3
	11	4
	9	3
	0	4
350		
5		
1384		

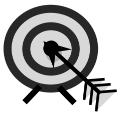
5. List the data values in the stem-and-leaf plot.

Stem	Leaf
5	0 1 4 8
6	2 6 7
7	1 4 5 6 6
8	2

CORRECT YOUR WORK. ANSWERS ARE AT THE END OF TOPIC 1.

Lesson 5: Continuous Numerical Data

You have defined stem plots and used them to display and organize data in the previous lesson.



In this lesson, you will:

- identify the steps in organizing continuous numerical data on a frequency table
- organize continuous numerical data on a frequency table.

You learnt something about continuous data in Lesson 1. Here again is the meaning of continuous data.

Continuous Numerical Data are data where every number on a scale has meaning. They are data which can take any value within a certain range.

As you have learnt continuous data are the result of measuring. So if collecting data involves measuring, then it is probably continuous numerical data. Most physical measurement can take decimal values and so are continuous data. This type of data will need to be grouped into classes so that it can be analysed.

Examples of continuous numerical data

(1) the volume of a bottle

(2) average numbers of people

(3) the width of a component

(4) temperature in a day

(5) time to produce an item

(6) heights in cm of the students in a class

To organize continuous numerical data in a frequency distribution table, we use the same approach as with the discrete numerical data.

Example 1

The ages of the students competing in an athletic meet are shown below.

13	14	11	14	16,	14
12	13	15	14	12	13
16	12	14	15	11	14
15	13	16	15	16	16

Display the result in a frequency distribution table.

Solution:

(i) List all possible ages in one column, the first row of the column having the lowest age, the last having the highest.
For the list above, these ages are 11, 12, 13, 14, 15 and 16.

(ii) Read through the list of scores. Each time a score occurs put a tally mark, which is a stroke (/) against the age. To make counting the scores easier the tally marks are grouped in fives (////), the fifth stroke being drawn diagonally across the first four.

(iii) When we have read through the list of ages, we count the tally marks for each age. This gives the frequency for each age.

(iv) Construct the frequency table to display the data.

Ages	**Tally Marks**	**Frequency**
11	*//*	2
12	*///*	3
13	*////*	4
14	*//// - /*	6
15	*////*	4
16	////	5
	Total:	24

This is a frequency table of individual ages. The total frequencies should always be checked to make sure it is the same as the number of original ages.

Example 2

The heights of the girls in the same year at a school were measured. The results are arranged in an array as follows.

155	156	156	156	157	157	157	157	157	158
158	158	158	159	159	159	159	159	159	159
159	159	160	160	160	160	160	160	161	161
161	161	161	161	161	161	162	162	162	162
162	162	162	163	163					

Organize the result in a frequency distribution table.

Solution:

(v) List all possible heights in one column, the first row of the column having the lowest height, the last having the highest.
For the list above, these heights are 155, 156, 157, 158, 159, 160, 161, 162, and 163.

(vi) Read through the list of scores. Each time a score occurs put a tally mark, which is a stroke (/) against the age. To make counting the scores easier the tally marks are grouped in fives (////), the fifth stroke being drawn diagonally across the first four.

(vii) When we have read through the list of heights, we count the tally marks for each age. This gives the frequency for each height.

(viii) Construct the frequency table to display the data.

Height (cm)	**Tally Marks**	**Frequency**
155	*I*	1
156	*III*	3
157	*~~IIII~~*	5
158	*IIII*	4
159	*~~IIII~~- IIII*	9
160	*~~IIII~~ - I*	6
161	*~~IIII~~- III*	8
162	*~~IIII~~- II*	7
163	*II*	2
	Total:	45

The frequency table usually is drawn without the tally mark. The table can have the value going down or across.

For example here is a frequency table from the tally table above.

Height (cm)	**Frequency**
155	1
156	3
157	5
158	4
159	9
160	6
161	8
162	7
163	2
	Total = 45

or

Heights(cm)	155	156	157	158	159	160	161	162	163
Frequency	1	3	5	4	9	6	8	7	2

If the data collected is big, the data needs to be grouped into classes so that it can be analysed. More of these will be discussed on the next lessons.

NOW DO PRACTICE EXERCISE 5

Practice exercise 5

1. Here are the ages of the players in the school orchestra.

12	12	12	13	13	13	13	13
13	13	13	14	14	14	14	15
15	15	15	15	15	15	15	16
16	16	16	16	18	18		

Show the information in a frequency distribution table.

2. The ages of audience members at a rap concert are shown below.

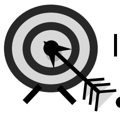
12	14	14	14	15	14	14	16,
11	14	15	15	12	12	11	13
14	16	14	14	13	13	14	15

Display the results in frequency distribution table.

CORRECT YOUR WORK. ANSWERS ARE AT THE END OF TOPIC 1.

Lesson 6: Grouped Frequency

You have described continuous numerical data and identify the steps to organize them on a frequency table in the previous lesson.



In this lesson, you will:

- described grouped frequency distribution table
- define a class, a class interval, class boundaries, the class size, and the class midpoint
- identify the steps involved in drawing a grouped frequency distribution table.
- draw a grouped frequency distribution table.

So far, you have learnt to construct frequency tables, giving a frequency for every individual score. However, if the scores are spread over a large range it is less time-consuming to just give the frequency of a group of scores.

Suppose you are asked to construct a frequency table of the entrance test scores of 120 Grade 9 students at FODE, what is the best thing to do?

In such cases where you are faced with lots of figures many of which will be the same, the best thing to do is to group them into smaller groups. Each group contains more than one score value, called the **class interval**. This class interval contains the number of score value. Let us look at how this is done by studying the example below.

Study the test scores of 40 students.

84	77	76	85	76	71	85	94	83	86
88	95	92	74	75	82	89	70	78	87
86	96	72	75	80	90	86	81	89	92
92	73	80	83	84	87	91	88	75	85

Notice that we only have the scores of 40 students here, but the method of dealing with the scores of 120 students in a similar problem is exactly the same.

Here are the steps to get the numbers we need to construct the frequency distribution table.

Step 1: Compute the range. This is the difference between the highest score and the lowest score. In the given data, the Highest score is 96 and the Lowest score is 70.

Hence, Range = 96 – 70

= 26

Step 2: Determine the class size.

Class size is the number of scores to be included in a sub-group or classes. First, we choose the number of sub-groups or classes. The number of classes formed is usually between 10 and 20.
Supposed we use 10 for our example. Then the class size is determined by dividing the range by the number of required classes.

$$\text{Class size} = \frac{\text{Range}}{\text{number of classes}}$$
$$= \frac{26}{10}$$
$$= 2.6$$

This indicates that each class or sub-group may have either 2 or 3 scores. Let us take 3.

Step 3: Organize the Class intervals or classes.

See to it that the lowest interval begins with a number that is a multiple of interval class size. Since the lowest score is 70, and the class size is 3, the lowest interval would begin with 69 and end at 71. These are the interval limits. Take note that the upper and lower limits (the exact or real limits) here are 68.5 and 71.5 respectively. These are sometimes referred to as **class boundaries.** To picture these limits, see illustration Figure 1.1 below.

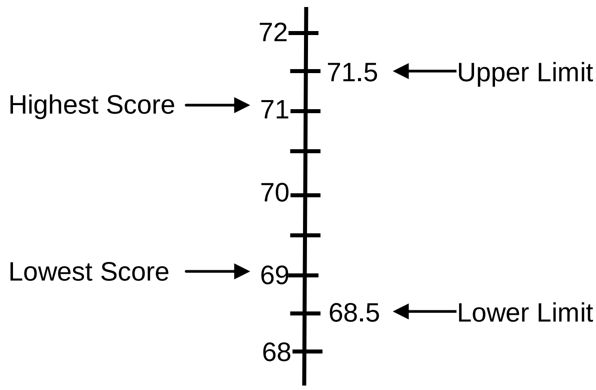


Figure 1.1: The vertical line showing the exact upper and lower limits.

After deciding upon the limits of the first class interval category, determine the rest of the intervals by increasing each interval limits by 3 until you reach the class 96-98 which contains the highest score in the distribution.

Let us start our first interval as 69-71. This includes 3 scores – 69, 70 and 71.
If we continue making the smaller groups, the next classes are 72-74, 75-77, 78- 80, and so on, until we reach the class containing the highest score which is 96 – 98.

Step 4: Tally each score to the category of class interval it belongs to.

Class Intervals	**Tally marks**
69-71	//
72-74	///
75-77	~~////~~ - /
78-80	///
81-83	////
84-86	~~////~~ - ///
87-89	~~////~~ - /
90-92	~~////~~
93-95	//
96-98	/

Step 5: Count the tally column and summarize it under column (f). Then add your frequency which is the total number of cases (N).

Class Intervals	**Tally marks**	**Frequency (f)**
69-71	//	2
72-74	///	3
75-77	~~////~~ - /	6
78-80	///	3
81-83	////	4
84-86	~~////~~ - ///	8
87-89	~~////~~ - /	6
90-92	~~////~~	5
93-95	//	2
96-98	/	1
		N = 40

Step 6: Compute the midpoint (M) for each class interval and put it under Column (M). You can obtain the midpoint by the formula below:

$$M = \frac{LS + HS}{2}$$

Where: M = the midpoint

LS = the lowest score in the class interval

HS= the highest score in the class interval

Class Intervals	Frequency (f)	Midpoint (M)
69-71	2	70
72-74	3	73
75-77	6	76
78-80	3	79
81-83	4	82
84-86	8	85
87-89	6	88
90-92	5	91
93-95	2	94
96-98	1	97
	N = 40	

Illustrative example for the first class interval:

$$M = \frac{69+71}{2} = \frac{140}{2} = 70$$

For the second class interval:

$$M = \frac{72+74}{2} = \frac{146}{2} = 73$$

Grouped Frequency Distribution is defined as the arrangement of the gathered data by categories plus their corresponding frequencies and class marks or midpoints. It has a class frequency containing the number of observations belonging to a class interval. Its class interval contains a grouping defined by the limits called the **lower** and **upper** limits. Between this lower and upper limits are the **class boundaries**.

NOW DO PRACTICE EXERCISE 6

Practice Exercise 6

1. Given are the following scores in a Chemistry test.

47	57	54	48	56	42	60	56
38	48	42	62	52	28	52	47
56	66	44	41	65	39	56	72
53	55	37	48	82	47	42	78
50	42	54	68	62	55	62	68

(a) Compute the range.

(b) Organize the class interval using a class size of 5. Your lowest class interval begins with 25 and end at 29.

(c) Make a frequency distribution table with the following feature columns.

Class Intervals	**Tally Marks**	**Frequency**	**Midpoints**
25 – 29			
30 – 34			
35 – 39			
40 – 44			
45 – 49			
50 – 54			
55 – 59			
60 – 64			
65 – 69			
70 – 74			
75 – 79			
80 - 85			

TOPIC 1: SUMMARY

- **Data** can be classified as Qualitative or Categorical (non-numerical) and Quantitative (numerical) data.
- **Qualitative or Categorical data** describes a characteristics or quality that cannot be counted.
- **Quantitative data** describes characteristics that has numerical value and can be counted or measured. They can be either discrete or continuous data.
- **Discrete data** are data that takes exact numerical values. It is often the result of counting.
- **Continuous data** are data measured on some scale and can take value within that scale. It is usually the result of measuring.
- A **Variable** is an object that is able to assume different values.
- **Organizing data** means arranging information so that it can be easily understood.
- **Continuous Numerical Data** are data where every number on a scale has meaning. They are data which can take any value within a certain range.
- **Grouped Frequency Distribution** is defined as the arrangement of the gathered data by categories plus their corresponding frequencies and class marks or midpoints.
- To construct a grouped frequency distribution table do the following steps;
  1) Compute the difference between the highest score and lowest score in the given set of data.
  2) Determine the class size. **Class size** is the number of scores to be included in a sub-group or classes.
  3) Organize the class intervals or classes.
  4) Tally each score to the class interval it belongs to.
  5) Count the tally column and summarize it under column (f). Then add your frequency which is the total number of cases.
  6) Compute the midpoint for each class interval. The **Midpoints or Class mark** of a class interval is the average of the lowest score and the highest score in the class interval. It is obtained by the formula:

$$\boxed{M = \frac{LS + HS}{2}}$$

ANSWERS TO PRACTICE EXERCISES 1-6

Practice Exercise 1

1.
   a) discrete
   b) continuous
   c) categorical
   d) discrete
   e) continuous
   f) continuous
   g) continuous
   h) categorical
   i) categorical
   j) continuous
   k) continuous
   l) categorical
   m) categorical
   n) discrete
   o) categorical

2. (a) discrete (b) colour

3. (a) continuous; continuous
   (b) discrete; continuous
   (c) continuous; categorical

Practice Exercise 2

1. (a)

Subjects	**Tally Marks**	**Frequency (f)**
English (E)	卌 – 卌 – 1	11
Science (S)	卌 – 卌	10
Mathematics (M)	卌 – 卌	10
Social Science (Ss)	卌	5
Personal Development (PD)	卌 - II	7
Design and Technology (DT)	卌 - III	8
	Total	51

(b) Favourite Subjects

(b) 7 students

(c) English

2. (a)

Name of Cars	Tally Marks	Frequency (f)
Ford (F)	//	2
Mazda (M)	~~////~~ – ~~////~~ - /	11
Suzuki (S)	~~////~~ – ~~////~~ - ///	13
Toyota (T)	~~////~~ – ~~////~~ - ~~////~~ – ~~////~~ – ////	24
Honda (H)	~~////~~ – ////	9
Nissan (N)	///	3
	Total	62

(b) 62 people

(c) Toyota

Practice Exercise 3

1. (a) Highest Score = 12 Lowest Score = 0
   (b)

Marks (Scores)	Tally Marks	Frequency
0	//	2
1	~~////~~ - /	6
2	~~////~~ - //	7
3	~~////~~ - //	7
4	~~////~~ - ///	8
5	~~////~~ -///	8
6	////	4
7	///	3
8	//	2
9	/	1
10	//	1
11		0
12	/	1
	Total	50

2.

Number of Goals (Scores)	Tally Marks	Frequency
0	///	3
1	~~////~~- /	6
2	~~////~~ - /	6
3	///	3
4	///	3
5	//	2
6	/	1
	Total:	24

3. (a) H.S. = 12; L.S. = 2

(b)

Marks (Scores)	**Tally Marks**	**Frequency**
2	///	3
3	~~////~~	5
4	~~////~~ - ////	9
5	~~////~~ - ~~////~~ - //	12
6	~~////~~ - ~~////~~ - ////	14
7	~~////~~ - ~~////~~ - ~~////~~ - //	17
8	~~////~~ - ~~////~~ - ~~////~~	15
9	~~////~~ - ~~////~~ - /	11
10	~~////~~ - ///	8
11	////	4
12	//	2
	Total	100

Practice Exercise 4

1.

Stem	**Leaf**
3	4 **8 3 1 9 3**
4	9 1 **9**
5	**7 9 5 1 3 8**
6	**1 8 3 1 0**

2.

Stem	**Leaf**
3	**7 9 5 2**
4	**0 4 5 1 9 7 8**
5	**9 2 8 9 2**
6	**6 2 8 2**

3. (a) 25

(b) 3 and 36

(c) 5

(d) 4

4.

Score	Stem	Leaf
39	**3**	**9**
27	**2**	**7**
125	**12**	**5**
83	8	3
114	11	4
93	9	3
4	0	4
350	**35**	**0**
5	**0**	**5**
1384	**138**	**4**

5. 50 51 54 58 62 66 67 71 74 75 76 76 82

Practice Exercise 5

1.

Age	Frequency
12	3
13	8
14	4
15	8
16	5
17	0
18	2
	Total = 30

2.

Age	Frequency
11	2
12	3
13	3
14	10
15	4
16	2
	Total = 24

Practice Exercise 6

1. (a) Range = HS – LS
= 82 – 28
= 54

(b) and (c)

Class Intervals	**Tally Marks**	**Frequency**	**Midpoints**
25 – 29	*I*	1	27
30 – 34		0	32
35 – 39	*III*	3	37
40 – 44	~~*IIII*~~ - *I*	6	42
45 – 49	~~*IIII*~~ - *I*	6	47
50 – 54	~~*IIII*~~ - *I*	6	52
55 – 59	~~*IIII*~~ - *II*	7	57
60 – 64	*IIII*	4	62
65 – 69	*IIII*	4	67
70 – 74	*I*	1	72
75 – 79	*I*	1	77
80 - 84	*I*	1	82

END OF TOPIC 1

TOPIC 2

PRESENTATION OF DATA ON GRAPHS

Lesson 7:	**Pictographs**
Lesson 8:	**Bar Graphs**
Lesson 9:	**Compound Graphs**
Lesson 10:	**Histograms and Frequency Polygons**
Lesson 11:	**Cummulative Frequency Tables and Graphs**
Lesson 12:	**Relative Frequency**

TOPIC 2: PRESENTATION OF DATA ON GRAPHS

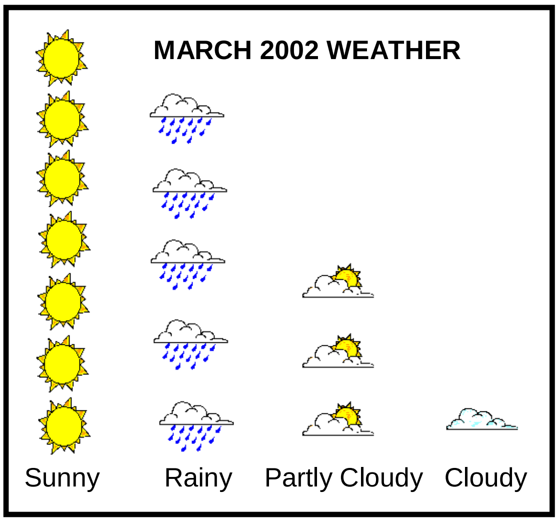
Introduction

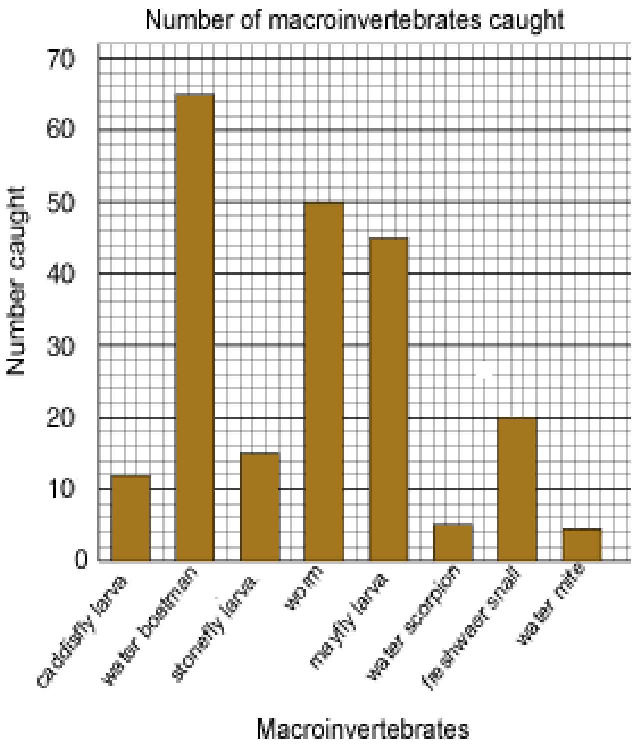
When frequency tables or distribution are drawn up the intension is that the table should tell us what sort of data and spread of data we have.

Some people find it easy enough to see these characteristics from the table but for many people is still a mass of numbers, so an alternative simpler method of presentation is required.

As we are trying to picture what our data is like we use pictures or pictorial representations of data using graphs.

Graphs are really pictures of statistical information. Here are some of them.

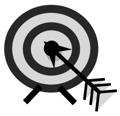




In this topic, you will further extend your knowledge and skills in presenting and displaying data using the different types of statistical graphs like **pictographs, bar graphs, compound graphs** in the first three lessons. Then you will look at the presentation of data using the **histogram**, **frequency polygon, cumulative frequency curves** known as “**ogives**” and **relative frequency polygon**.

Lesson 7: Pictographs

Welcome to the first lesson of Unit 3 Topic 2. You have already learnt something about pictograph in your Grade 7 and 8 Mathematics courses.



In this lesson, you will:

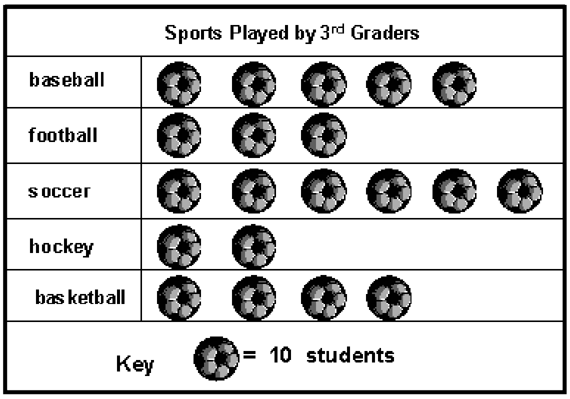
- revise and define pictograph
- present data on pictograph

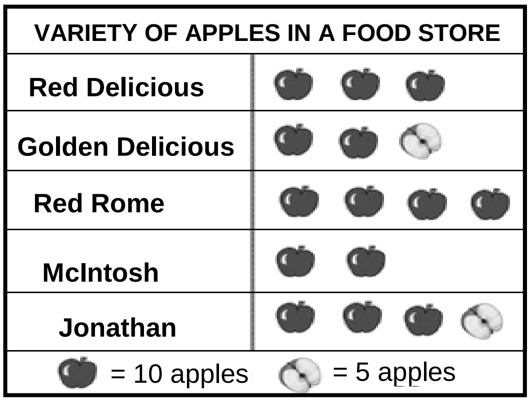
Here is the definition of pictograph again if you don't remember.

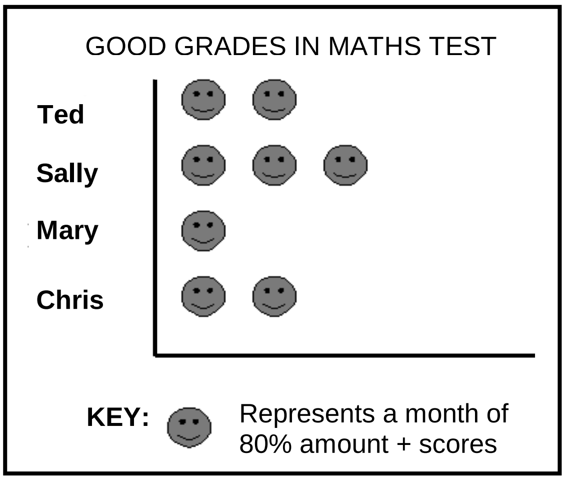
A Pictograph is a graph which uses pictures or symbols to represent statistical information or data. It is a way of representing data using symbolic figures to match the frequencies of different kinds of data.

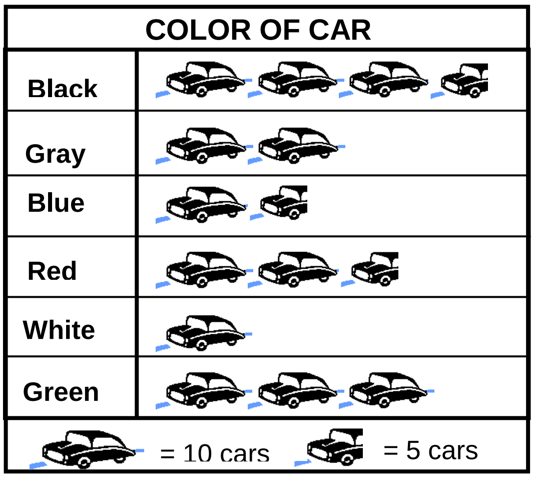
Pictographs can be found in the works of many ancient cultures in papyrus, wood cloth, pottery and painted on walls. Sometimes pictographs are used to describe pictures or symbols carved or chipped in rock (petroglyphs).

Pictographs are pictures or picture-like symbols that represent an idea or tell a story. Here are some examples of pictographs.

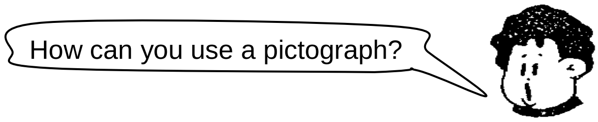








Sometimes pictographs are called **pictograms or picture graphs.**



You can use a pictograph to represent different amounts of data.

A pictograph takes the form of a bar graph.

The key for a pictograph tells the number that each picture or symbol represents.

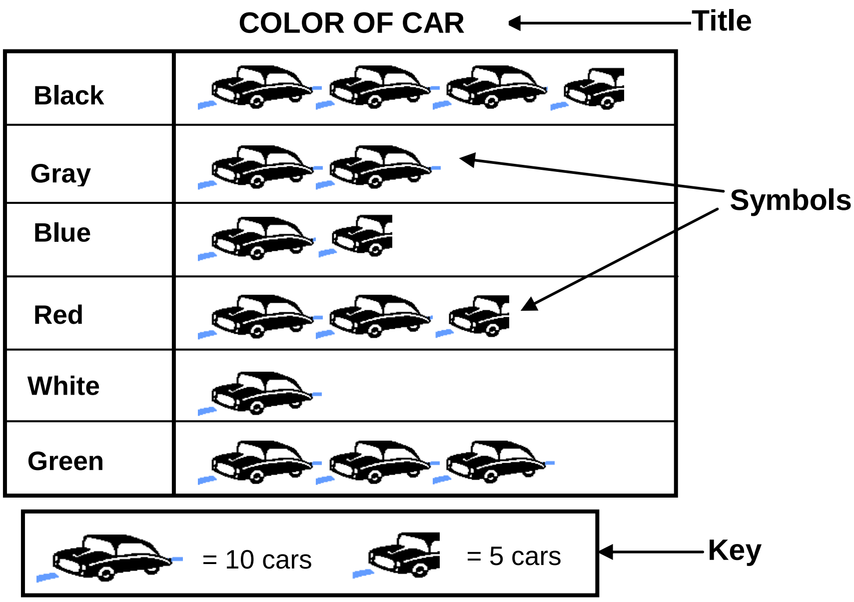
Using a pictograph has some advantages.

1. A pictograph is easy to read.
2. They show trends in data clearly.
3. They are fun to use.

But there are also disadvantages.

1. It may be difficult to find a symbol or picture to represent the data.
2. The key can be confusing to read.
3. A pictograph can be difficult to make.

Here is a pictograph which we will use to describe the main points about a pictograph.



REMEMBER

A pictogram must have: **(1)** a **Title** to explain what the graph is about.
(2) a **Key** to show what each **symbol** stands for.

We can use the information from the pictograph, to answer questions.

For example:

(a) In the pictograph what is the value of a whole car?

Answer: Looking at the Key, one whole car represents 10 cars.

(b) What color of car is most popular?

Answer: You will see that in the pictograph black has the symbol:

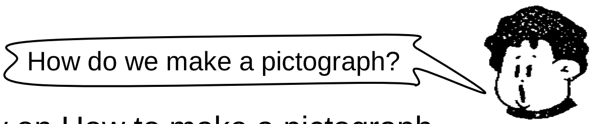
10 + 10 + 10 + 5 = 35 cars

Therefore, black is the most popular color.

(c) How many cars are red?

Answer: Since the symbol  represents 10 cars and  represents 5 cars.

Therefore, the number of red cars is 25.



Follow the steps listed below on How to make a pictograph.

How to make a pictograph.

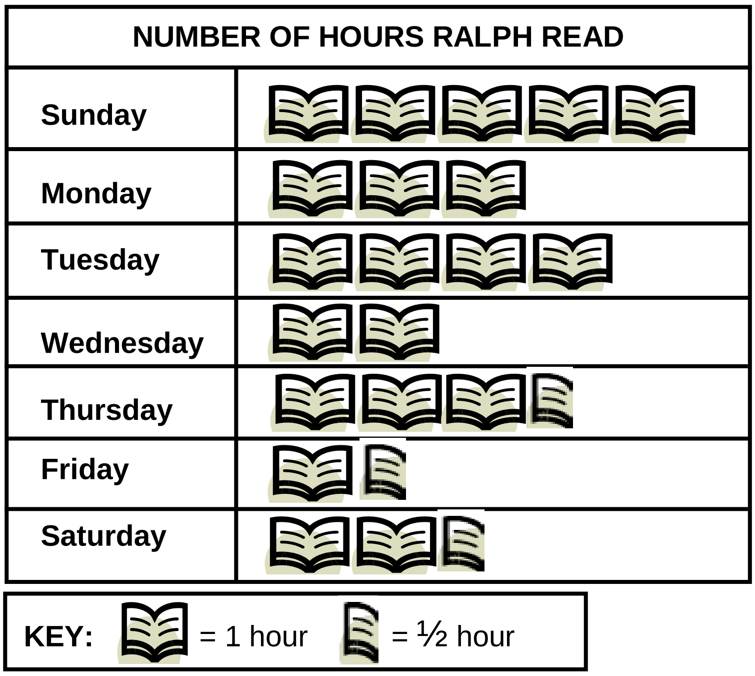
1. List each category.
2. If necessary, round off the data to the nearest whole numbers.
3. Choose a picture or symbol that can represent the number in each category.
4. Choose a key.
5. Draw pictures to represent the number in each category.
6. Label the pictograph. Write the title and the key.

Now let us use the table below to make our pictograph.

NUMBER OF HOURS RALPH READS	
Sunday	5
Monday	3
Tuesday	4
Wednesday	2
Thursday	3½
Friday	1½
Saturday	2½

To show the data in a pictograph, we use the symbol  to represent 1 hour.

The pictograph looks like this:



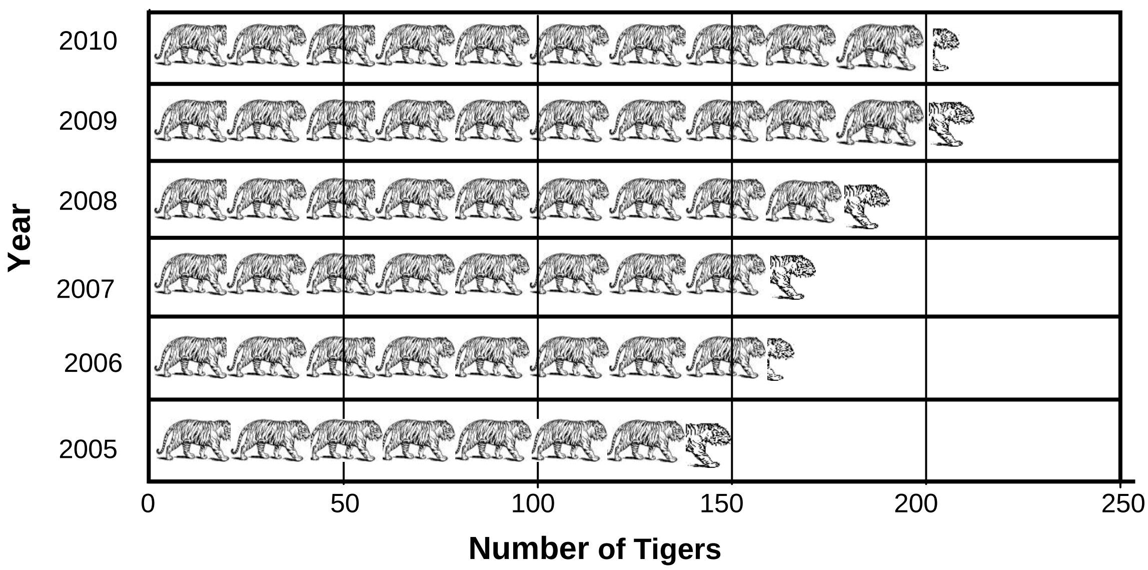
Here is another example.

This table shows a data on the number of tigers living in a game reserve in different years.

Year	2005	2006	2007	2008	2009	2010
Number of Tigers	150	165	172	190	218	205

To show the data with a pictograph, we need to choose a scale because the numbers are large. If we use one tiger symbol to represents 20 tigers, the pictograph looks like this:

TIGERS LIVING IN A GAME RESERVE



Note that if the number of tigers does not divide by 20, you need to draw part or portion of the tiger.

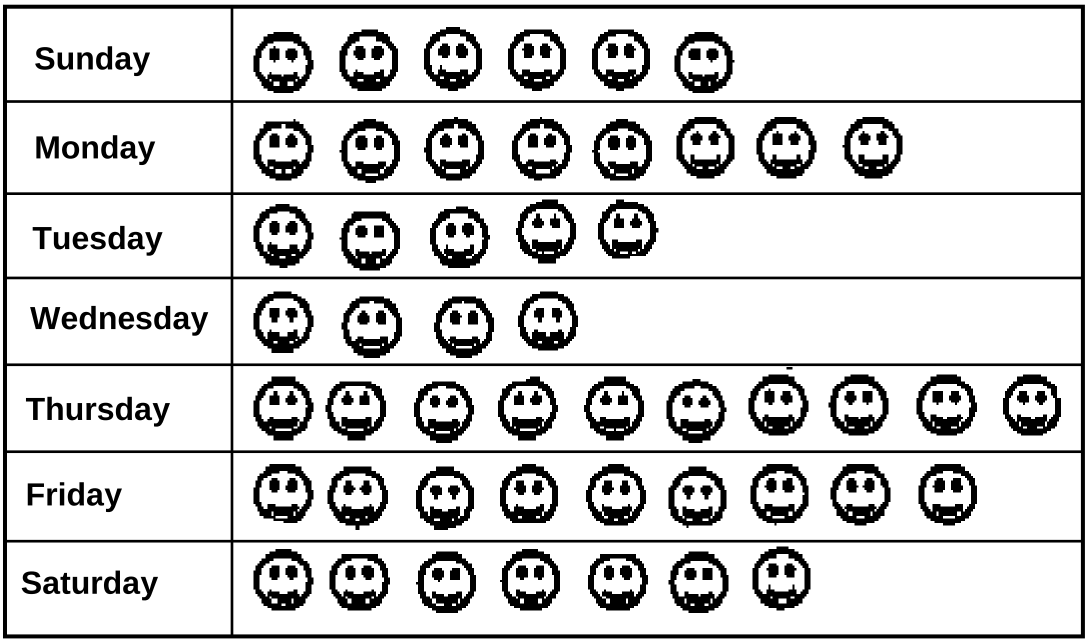
Drawing the same symbol many times can be very boring. So, you have to select very simple symbols for pictographs.

In making your pictograph, remember you have to choose a picture or symbol to represent your data. Make sure your key explains how much each picture or symbol is worth.

NOW DO PRACTICE EXERCISE 7

Practice Exercise 7

1. The pictograph given below expresses the number of persons who travelled from Central Province to NCD by PMV on each day of a week.



KEY:  = 50 persons

From the pictograph gather the information and answer the following questions:

(a) How many travellers travelled each day of the week from Central Province to NCD?

(b) On which day was there a maximum rush for the PMV?

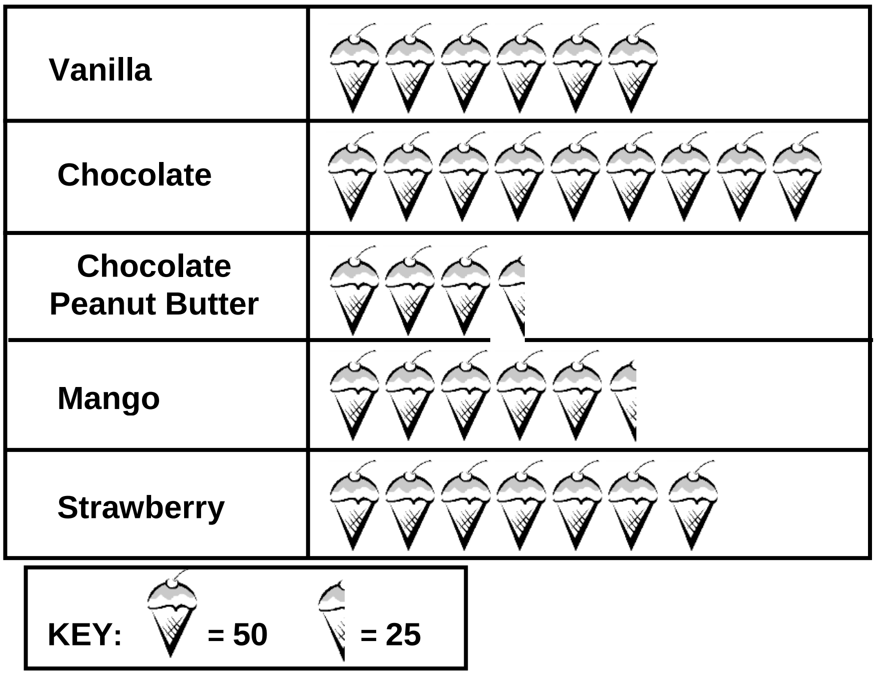
(c) How many travellers travelled during the week?

(d) On which day was there a minimum rush for the PMV?

(e) Find the difference between the number of travellers who travelled in maximum and minimum numbers.

2. The pictograph shows the number of ice cream cones sold during the days of a week from a shop. Give the following information regarding sale of toys.

ICE CREAM CONES SOLD



(a) How many chocolate ice cream cones were sold?

(b) How many strawberry ice cream cones were sold?

(c) Which type of ice cream was sold the least?

(d) Did more people buy vanilla than mango ice cream cones?

---

3. Shawn asked his friends what hobbies they had. His results are recorded in a table as shown.

Hobby	Frequency
Computer Games	12
Football	18
Music	6
Others	9

(a) How many people chose computer games as one of their hobbies?

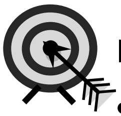
(b) Draw a pictograph to show Shawn’s results.

Use the symbol  to represent 3 persons.

CORRECT YOUR WORK. ANSWERS ARE AT THE END OF TOPIC 2

Lesson 8: Bar Graphs

You learnt to present data on pictographs or pictograms in the previous lesson.



In this lesson, you will:

- define bar graph
- present data on a bar graph.

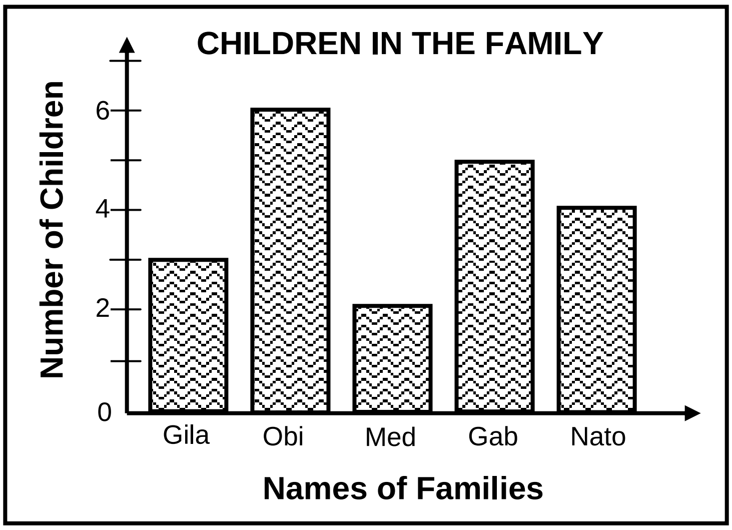
You have already learnt about bar graphs in your grade 7 and 8 Mathematics courses.

Here again is the definition of bar graphs.

Bar Graphs are graphs which use parallel bars with equal width to show statistical data. The length of the bars is drawn proportional to the quantities they represent. The bars are drawn horizontally or vertically.
Bar graphs are used to show how quantities compare in size.

When the bars are drawn vertically, the bar graph is called a **column graph** or **vertical bar graph**.

Here is an example of a column graph.



We can use the information in the column graph and interpret it to answer question such as:

Which family had the most children? **Answer: Obi's**

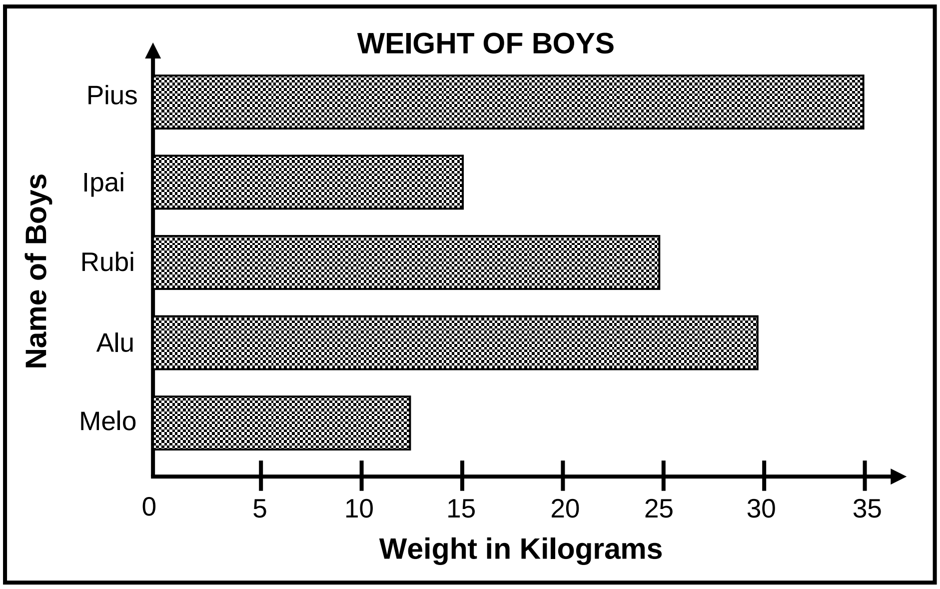
What was the least number of children in a family? **Answer: 2 children**

What is the average number of children per family?

(3 + 6 + 2 + 5 + 4 = 20; 20 ÷ 5 = 4) **Answer: 4 children**

When the bars are drawn horizontally, the bar graph is called a **horizontal bar graph**.

Here is an example of a Horizontal bar graph.



Let us answer the following questions using the information from the bar graph above. To read a graph like this we need to know the **scale** of the **horizontal axis**.

On the horizontal axis, one (1) centimetre represents 5 kilograms. Therefore, the scale is 1 cm : 5 kg.

For example:

Melo's bar is 2.5 cm, so 2.5 x 5 kg = 12.5 kg.

(a) List the boys in ascending order of their weights.

Melo	12.5 kg
Ipai	15 kg
Rubi	25 kg
Alu	30 kg
Pius	35 kg

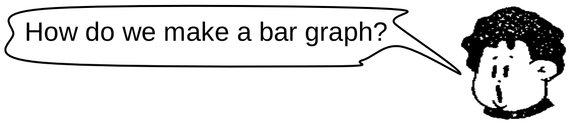
(b) What is the difference between the weights of the heaviest and the lightest boy?

Difference in weight = Wt. of heaviest boy – Wt. of lightest boy

= 35 – 12.5

= 22.5

Therefore, the difference in weight is 22.5 kg.



Remember, to make or draw a column or a horizontal bar graph, involves a lot of steps. Here are 4 steps to help you.

STEP 1 Work out the scale for each axis to determine the length of each axis and each bar using the information.

STEP 2 Draw the scaled axes, number the axes and label them.

STEP 3 Draw the bars. The bars should be of the same width and the spaces between them should be the same.

STEP 4 Give a brief title to the graph.

Example 1

Here is a table showing Paru's test result.

Subjects	**Percentage**
English	70%
Maths	95%
Science	65%
Commerce	85%
Social Science	80%

We will use the information to draw and make a horizontal bar graph.

The **subjects** will be shown on the vertical axis and the **percentages** will be shown on the horizontal axis.

STEP 1 Scale: there are 5 subjects. If we draw 5 bars (one for each subject) and we make each 0.5 cm wide and 0.2 cm space between the bars, we need about 5 x 0.5 + 5 x 0.2 = 2.5 + 1 = 3.5 cm or 4 cm length on the vertical axis.

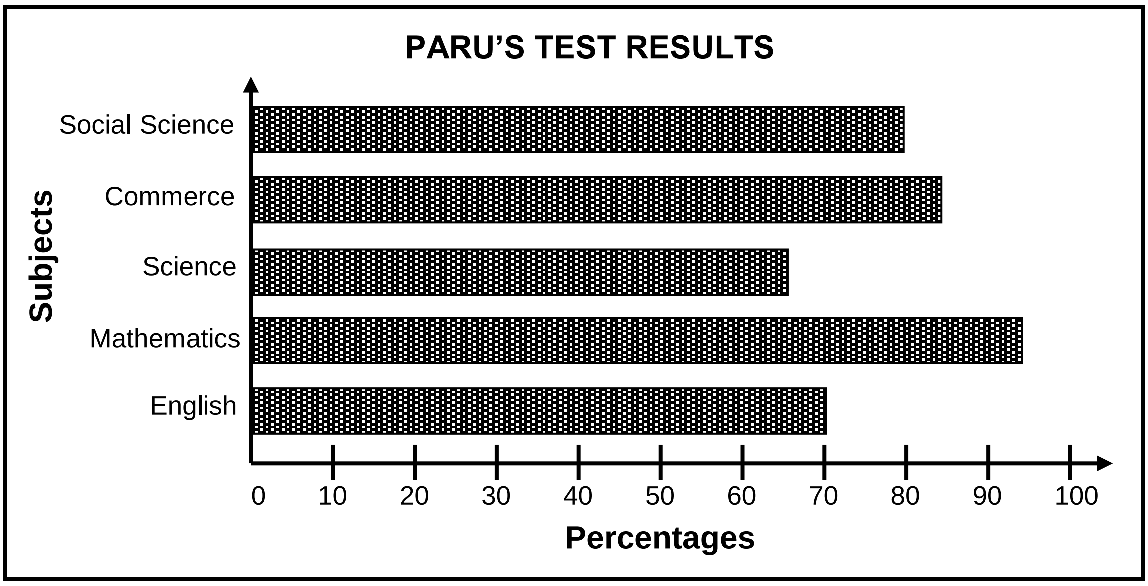
We need to show marks up to 100% because the highest mark of 95% is close to 100%. If we use 1 cm to represent 10%, we will need 100 ÷ 10 cm on the horizontal axis.

The scale for the horizontal axis is 1 cm : 10%. That is 1 cm represents 10%.

A suitable **title** for the graph would be " **Paru's Test Results**".

STEP 2 T0 STEP 4 Draw the graph. (See next page).

The graph would look like this.



Example 2

The table shows the information on pawpaw picked by five boys.

Name	Number of Pawpaw
Kasa	8
Kiki	24
Nelson	32
Charlie	24
Benua	16

Draw a column graph using the information.

To draw the column graph, we use the same steps we used to draw the horizontal bar graph.

The names of the boys will be on the horizontal axis.
The number of pawpaw will be on the vertical axis.

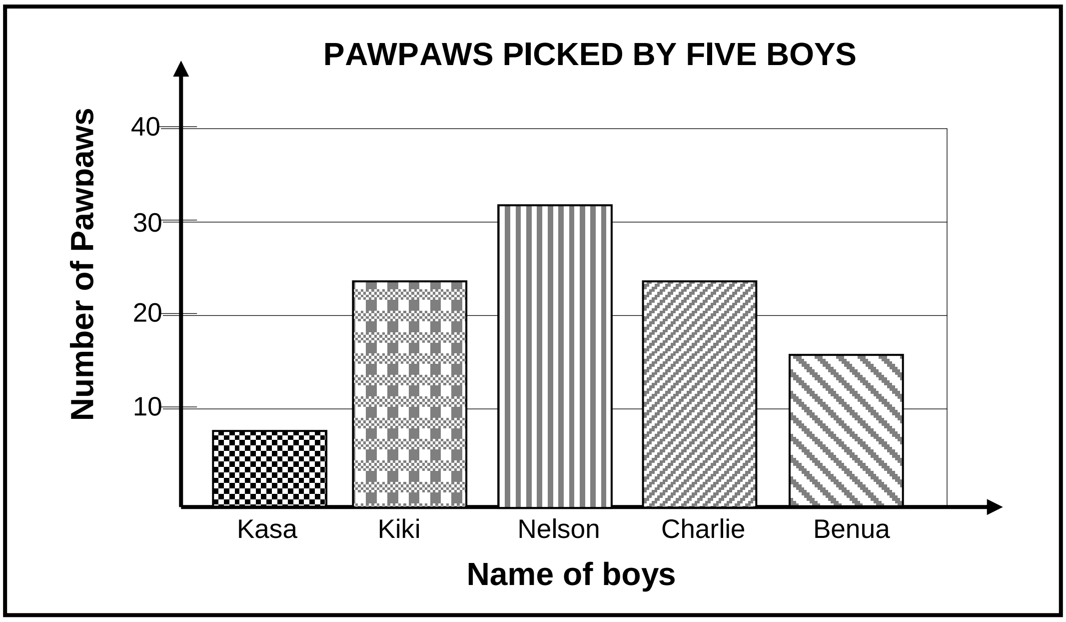
STEP 1 Scale: There are 5 boys, so we need about 1cm x 5 = 5 cm in length for the horizontal axis.

The highest number of pawpaws is 32 and the numbers are multiples of 8. So 32 ÷ 8 = 4 cm will be the height required.

The scale for the vertical axis is 1 cm:10 pawpaw. That is 1 cm represents 10 pawpaws

STEP2 – 4 Draw the bar graph. (See next page)

The graph would look like this:



Remember:

A bar graph is useful for comparing facts. The bars provide a visual display for comparing quantities in different categories (groups). Bar graphs help us to see relationships quickly. Each part of a bar graph has a purpose.

For example:

(a) The title tells us what the graph is all about.

(b) The labels tell us what kinds of facts are listed.

(c) The bars or rectangles show the facts.

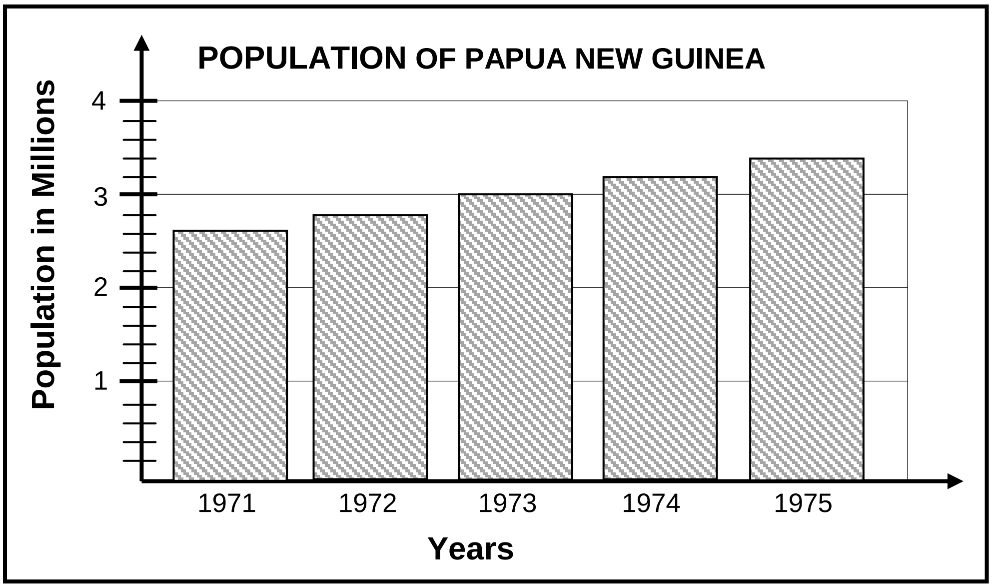
(d) The grid lines are used to create the scale

(e) Each bar shows a quantity for a particular category or group.

NOW DO PRACTICE EXERCISE 8

Practice Exercise 8

1. Here is a graph showing the population of Papua New Guinea from 1971 to 1975.



Answer the following questions using the information in the graph.

(a) What was the population in 1973?

(b) What was the increase in population from 1971 to 1975?

(c) Was there likely to be a population increase in 1976?

(d) Give a possible reason for the increase in population from 1971 to 1975.

2. A survey of students favourite after- school activities was conducted at a school. The table below shows the results of this survey.

STUDENT'S FAVOURITE AFTER-SCHOOL ACTIVITIES	
Activity	**Number of students**
Play sports	45
Talk on Phone	53
Visit with friends	99
Earn Money	44
Chat online	66
School Clubs	22
Watch TV	37

(a) i. Which after-school activity do students like the most?

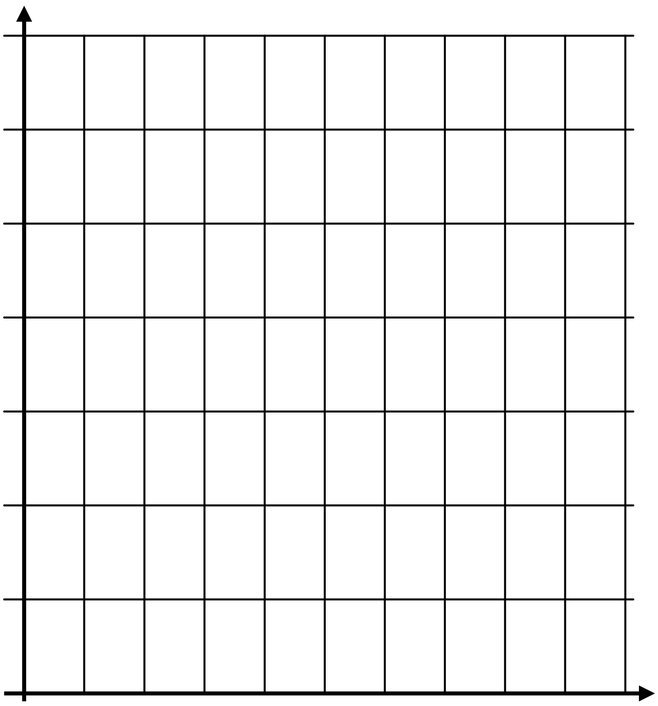
ii. Which after-school activity do students like the least?

iii. How many students like to talk on the phone?

iv. How many students like to earn money?

v. List the categories in the table from greatest to least?

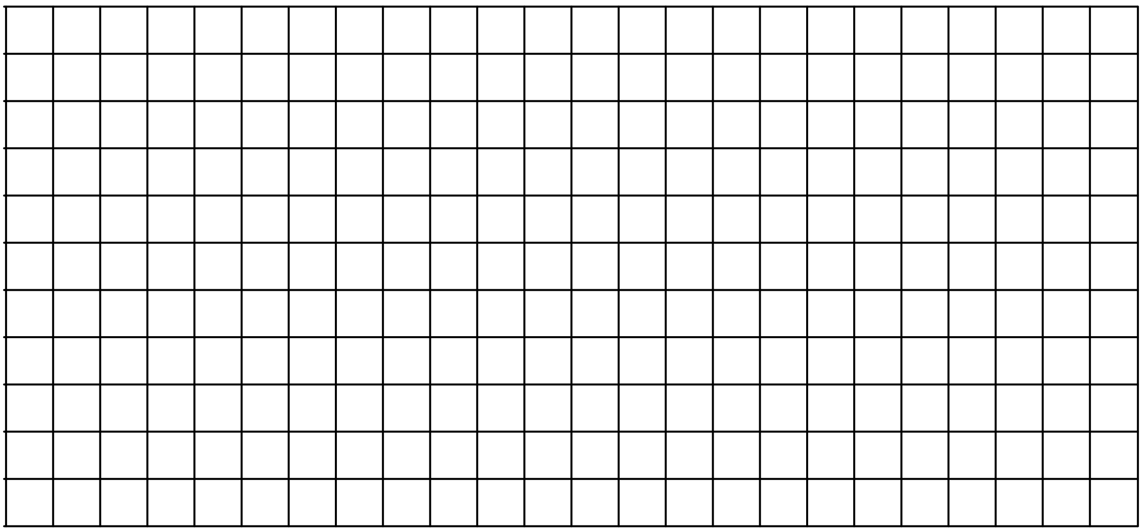
(b) Draw a horizontal bar graph showing the information from the above table in the grid below. Use 1 division = 20 students along the horizontal axis.



3. Here is a table showing how John planned to use his salary of K400.

Items	Amount
Food	140
Rent	80
Transport	60
Savings	40
Clothing	30
Services	30
Entertainment	20

(a) Draw a horizontal bar graph in the grid below using the information from the table above. Use 1 division = 20 kina along the horizontal axis.



(b) Answer the following questions using the information presented in the graph.

i. On what item will John spend most of his money?

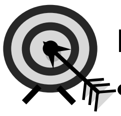
ii. On which item will the least amount of money be spent?

iii. What percentage of the pay did John spend on rent?

CORRECT YOUR WORK. ANSWERS ARE AT THE END OF TOPIC 2

Lesson 9: Compound Graphs

You have revised and learnt how to present and interpret statistical data and information with a bar graph in the last lesson.



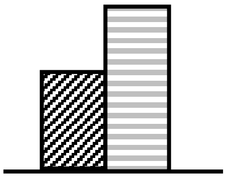
In this lesson, you will:

- identify and describe features of compound graphs
- present data on a compound graphs.

Earlier in your study of Grade 8, you learnt to identify different sets of information presented in a compound graph.

A compound graph is a special type of bar graph that compares two or more quantities simultaneously in one graph.

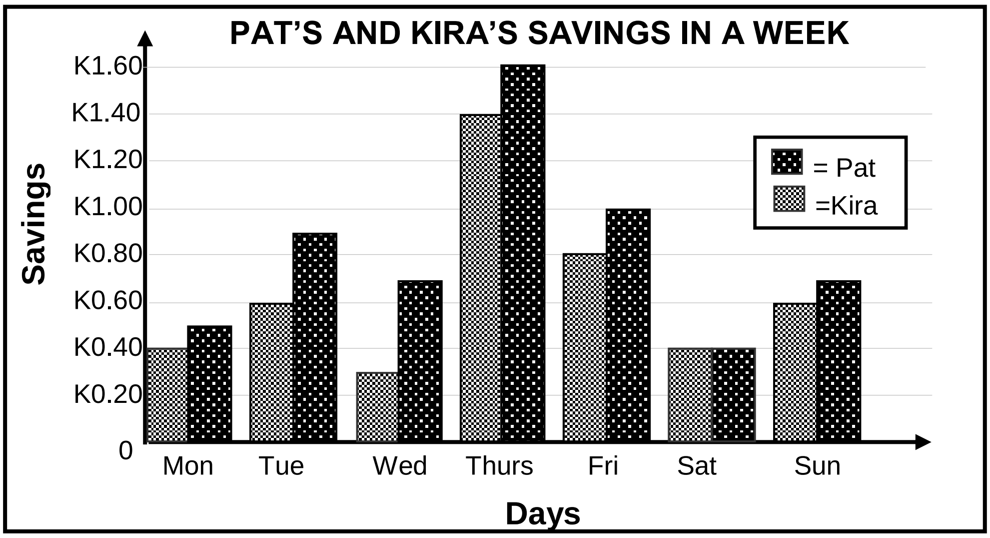
When a compound graph is drawn with different bars beside each other like this one below, it is called **a compound bar graph.**



We can use compound bar graph to show and compare data for two related items for the same period.

Example 1

Here is a compound bar graph to compare Pat's and Kira's savings.



Notice that the bars are beside each other and have different shading.

Look at the **key.** It explains the two types of bar.

This kind of bar graph is useful to compare two sets of information.

Now using the graph in the previous page, answer the following?

(a) What is the graph about?

Answer: Pat's and Kira's savings in a week

(b) At a glance, can you tell who is thriftier?

Answer: Yes, Pat.

(c) What is the total savings of each girl in a week?

Answer: Pat = K5.80 and Kira K4.50

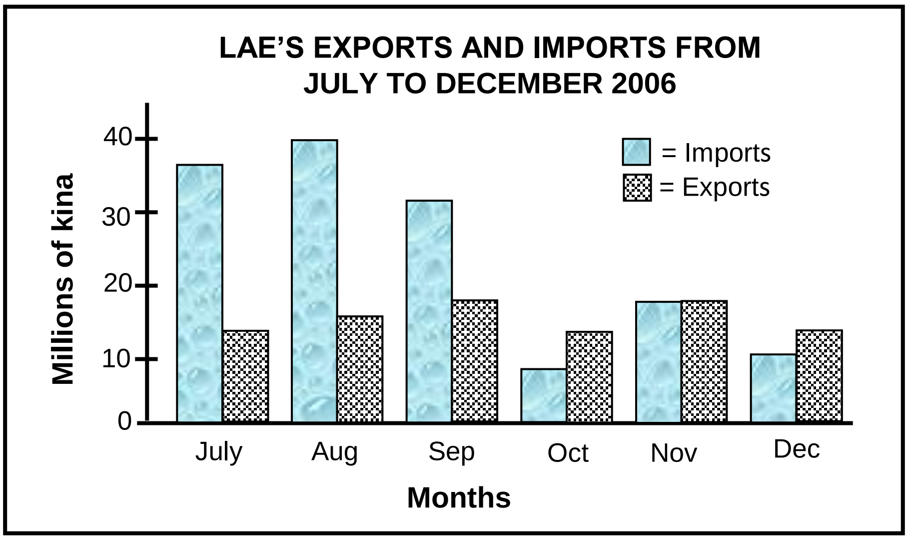
(d) What per cent of his allowance does Kira save in a week? How about Pat?

Answer: Kira = 56.25%

Pat = 72.5%

Example 2

Here is another example of compound bar graph which compares exports and imports through Lae from July to December 2005.



We can use this compound graph to compare exports and imports.

(a) In which months were exports equal to imports?

Answer: November

(b) When were exports greater than imports?

Answer: October and December

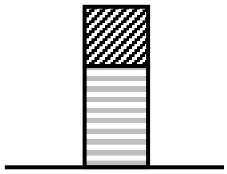
(c) When were exports less than imports?

Answer: July, August and September

(d) Were exports or imports greater over the six (6) months period

Answer: Imports were greater than exports.

Sometimes we draw different bars on top of each other like the one below.

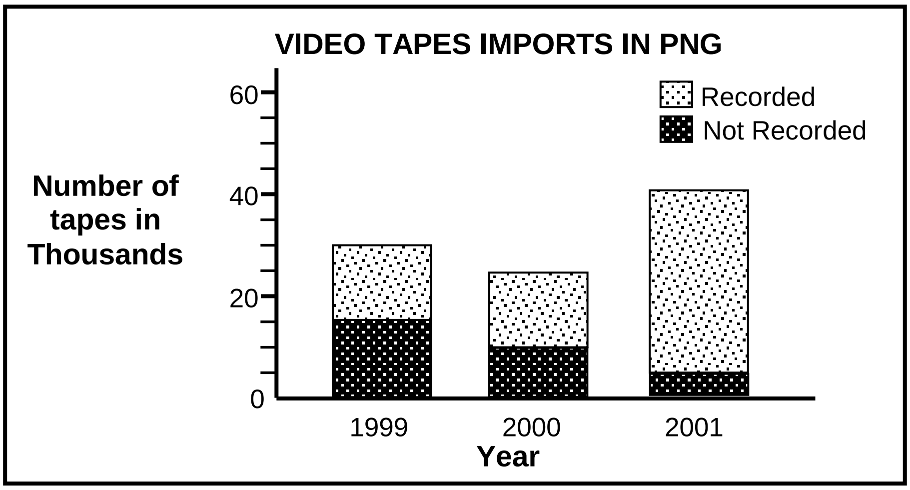


We call this type of compound graph a **stacked bar graph.**

A Stacked bar graph is a graph that is used to compare the parts to the whole. The bars are divided into categories or group. Each bar represents a total.

Example 3

Here is an example of a stacked bar graph which compares video tapes, recorded and not recorded.



Notice that the bars are stacked on top of each other.

The key explains the two types of shaded bars representing the two groups: the recorded tapes and Not recorded tapes.

The total height of each compound bar gives the total number of tapes imported. The height of each bar division gives the number of recorded and not recorded tapes imported.

To find the number of "not recorded" that is imported, subtract the number of recorded tapes from the total number of tapes imported for a particular compound bar.

For example:

(a) How many not recorded tapes were imported in 1999?

Solution:

Number of not recorded tapes = Total number of tapes – Number of recorded tapes
= 95 000 – 65 000
= 30 000

Therefore, the number of not recorded tapes in 1999 is **30 000.**

This type of compound bar graph enables

(a) Totals of the bars compared correctly.
For example, the number of tapes imported nearly doubled by 2001.

(b) A comparison of the same type of bars.
For example, the number of recorded tapes imported decreased every year. Therefore the number of not recorded tapes increased every year from 1999 to 2001.

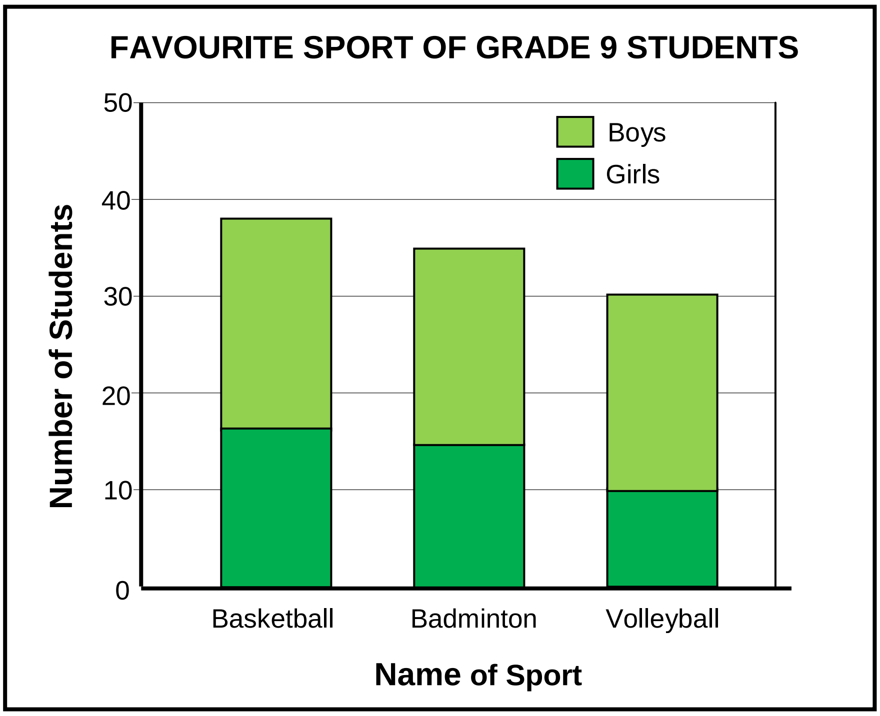
(c) Comparison of part bars.
For example, the number of recorded tapes was almost the same as the number of tapes not recorded in 1999.

Here is another example of stacked bar graphs.

In the following example, each bar of the stacked bar graph is divided into two categories or groups: boys and girls.

Each of the three bars represents a whole.

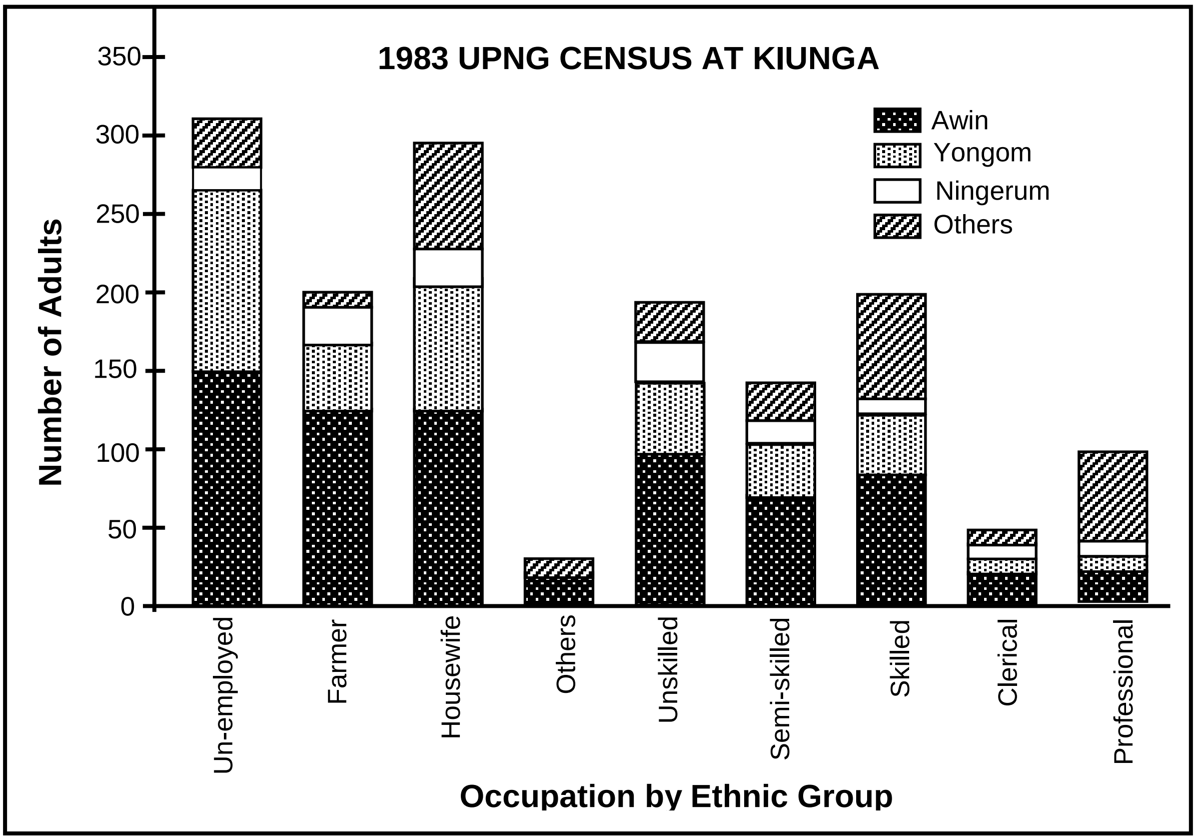
That is about 38 students who liked basketball, out of which 16 are girls.



NOW DO PRACTICE EXERCISE 9

Practice Exercise 9

1. Here is another graph showing the 1993 UPNG census at Kiunga.



Answer the following questions using the information in the above graph.

(a) The largest group is the unemployed. The second largest group is the ____________.

(b) The largest ethnic group is the ___________.

(c) There are __________ farmers than unskilled workers in Kiunga.

(d) Estimate the total clerical workers.

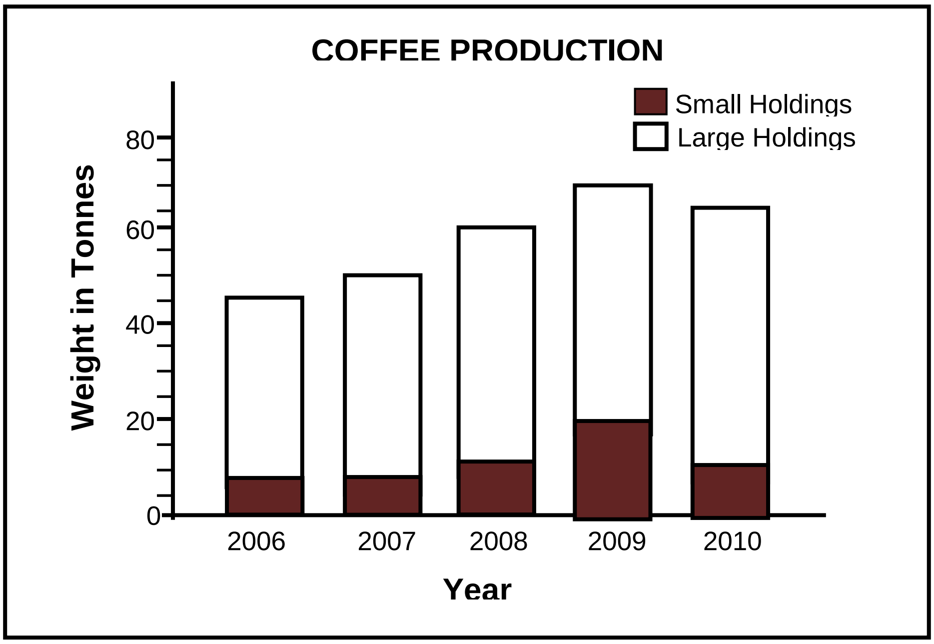
(e) Give the three main occupation of the Ningerum.

(f) Which statement below is true?

i. The majority of professional people do not come from the Awin, Yongom and Ningerum tribes.

ii. The majority of unskilled and semi-skilled workers belong to the Awin tribes.

2. Here is a stacked bar graph showing coffee produced from 2006 to 2010 in PNG.



Answer the following questions about the graph.

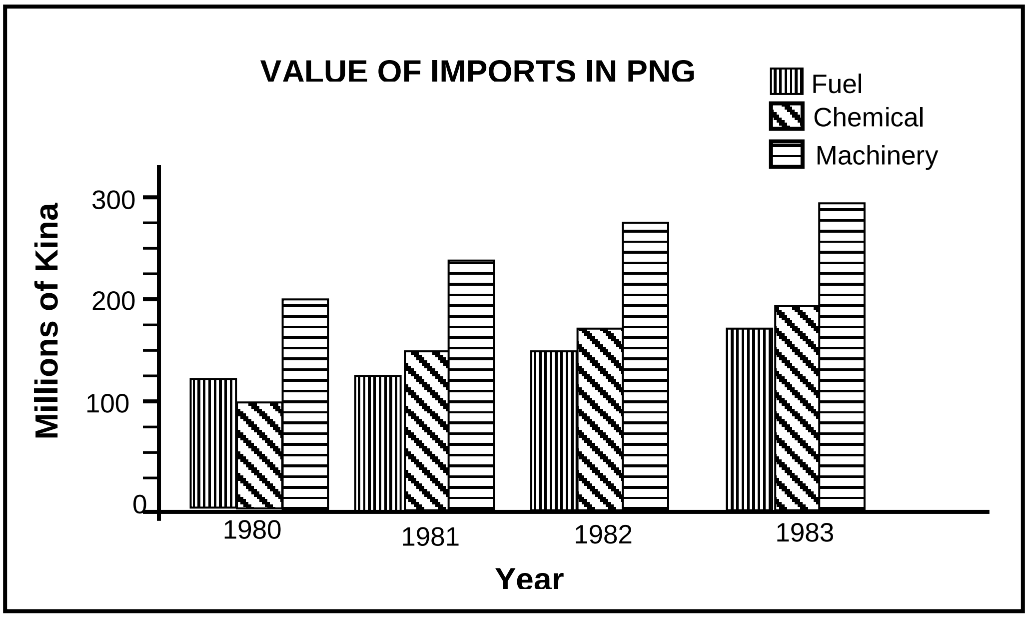
(a) Which bars represent the large coffee holdings?

(b) For how many consecutive years was the total coffee production increasing?

(c) In which two consecutive years did the production in small holdings remain the same?

(d) What year did PNG experience the first decrease in total coffee production?

3. Here is a compound bar graph showing the value of imports into PNG from 1980 to 1983.



Answer the following questions using the information on the compound graph.

(a) What is the main import into PNG?

(b) Estimate the total value of imports in 1982.

(c) Between 1980 and 1983, have chemical imports increased, decreased or remained about the same?

(d) Did the total value of imports increase or decrease between 1980 and 1983?

4. The table below shows the maximum marks scored by Grade 7, 8 and 9 students in Mathematics.

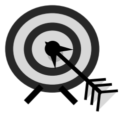
Grade	Girls	Boys
7	20	19
8	15	20
9	10	30

Draw a stacked bar graph on the box below showing the information above.

CORRECT YOUR WORK. ANSWERS ARE AT THE END OF TOPIC 2

Lesson 10: Histograms and Frequency Polygons

In your Grade 7 and 8 Mathematics you learnt how to present statistical data in a frequency table.



In this lesson, you will:

- define and identify features of a histogram and a frequency polygon
- identify the steps in making a histogram and frequency polygon
- draw a histogram and frequency polygon for a set of data

A convenient way or method of representing a frequency distribution graphically is by means of a frequency histogram.

You learnt something about histogram and frequency polygon in your study of Grade 7 and 8 Mathematics.

Let us revise the definition of a histogram and a frequency polygon.

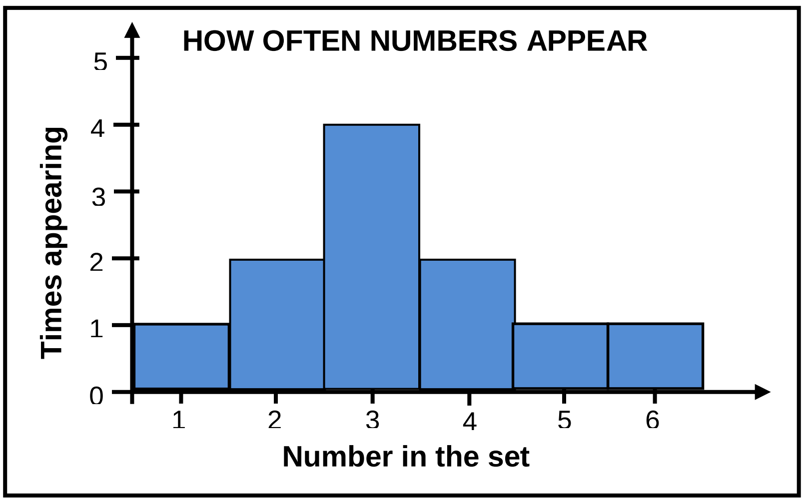
The histogram is a special type of bar graph where the bars are always vertical and are placed next to each other without gaps. The values of the variables or the scores are placed in the horizontal axis and the frequency of the variables or the scores on the vertical axis.

For example

Let us graph the following group of numbers below according to how often they appear.

1, 2, 2, 3, 3, 3, 3, 4, 4, 4, 5, 6

We can graph them like this.



The histogram is easy to make and gives us some useful information about the set. For example, the graph's highest point or peak is at 3, which is also the median and the mode of the set of numbers. The mean of the set of numbers is 3.27 which is also not far from the peak.

Example 2

Here is a frequency distribution table of the ages of a group of people that can be used to draw a Histogram.

Age Group	**Frequency**
1 - 10	1
11 - 20	3
21- 30	6
31 - 40	4
41- 50	2
Total	16

A histogram for a grouped distribution can be drawn by using the midpoints of the class intervals as centres of the bars.

To draw a histogram we need to work out the mid points or class centres of the age group.

Recalling the formula for midpoint or class centre, let us work out the class centre. Add the end points of each class interval and divide by 2.

For example

The class centre for the interval 1 – 10 is $\frac{1 + 10}{2} = 5.5$

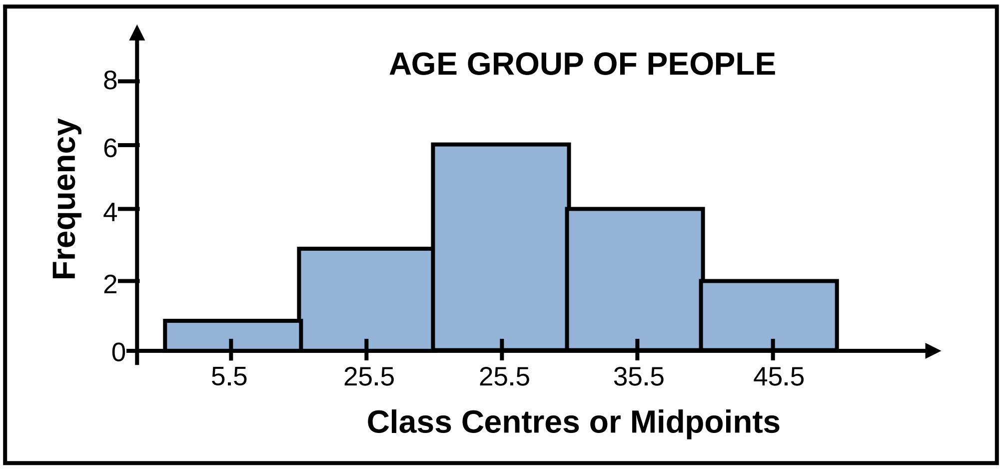
The class centre for the interval 11 – 20 is $\frac{11 + 20}{2} = 15.5$ and so on.

Here is the same table from the previous page showing the class centres of each class.

Class Intervals	**Class centre**	**Frequency**
1- 10	5.5	1
11 - 20	15.5	3
21- 30	25.5	6
31 - 40	35.5	4
41- 50	45.5	2
		Total = 16

Once the class centres are known a grouped frequency histogram can be drawn in the same way as the frequency histogram, but we plot the class centres of the class intervals on the horizontal axis rather than the original. On the next page is a histogram drawn from the distribution table above.

Here is the histogram.



The bars are centred about the ages they represent. They are the same width and are joined.

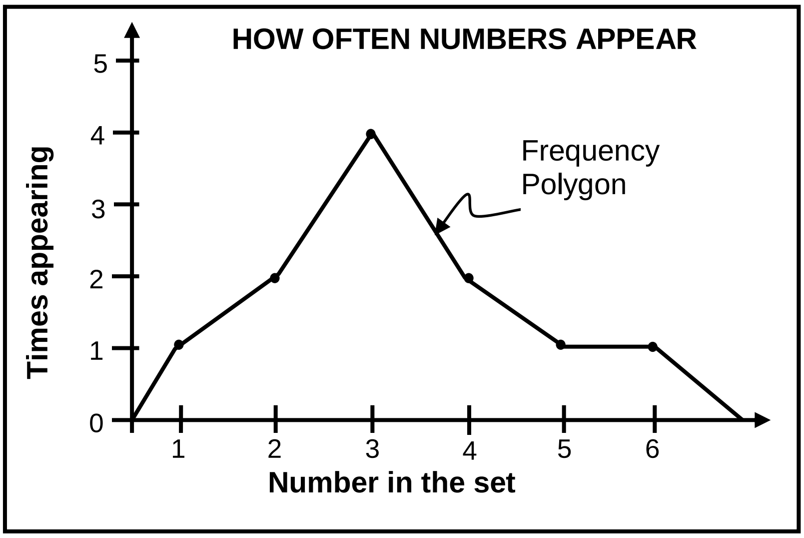
The area of each bar represents the frequency of each score. Hence, the total area of the histogram represents the total number of score.

Another way of representing frequency distribution graphically is the **Frequency Polygon.**

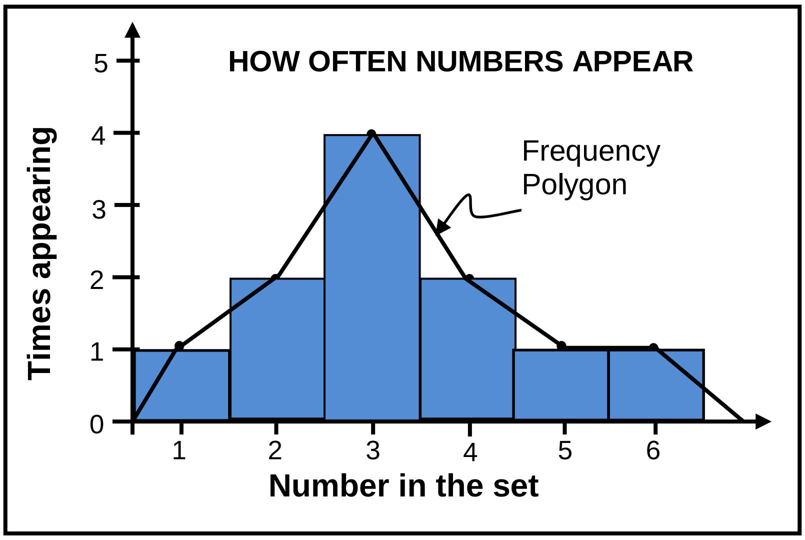
A Frequency polygon is a special kind of line graph which is drawn by joining all the midpoints of the top of the bars of a histogram.

For example

Here is the frequency polygon of the set of numbers according to how often they appear.

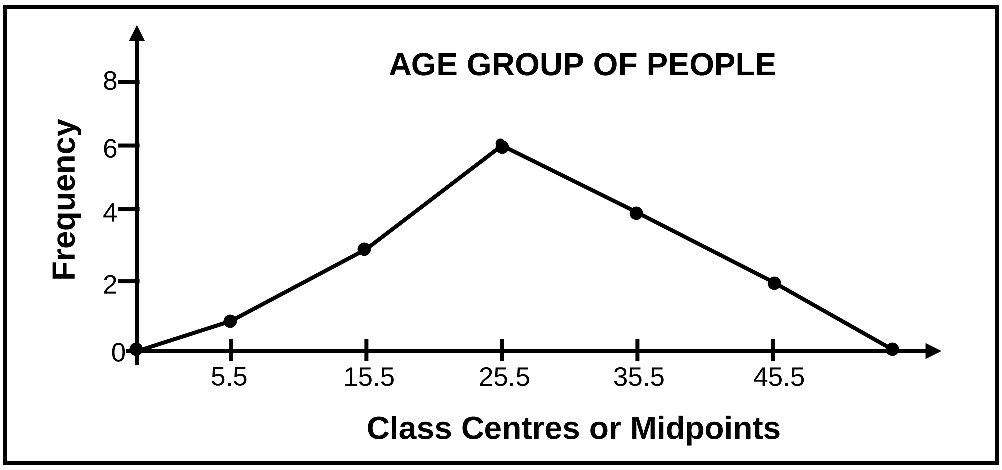


It is interesting to note that if a frequency histogram and polygon are drawn on the same axes, the polygon joins the midpoints of the top of each bar or column in the histogram. This can be seen in the diagram on the next page.



Example 3

Here is the frequency polygon of the age group of people.



Note that since the area under the polygon should be equal to the area of the histogram then the first and last points should be joined to the points on the horizontal axis where the next score would be found.

NOW DO PRACTICE EXERCISE 10

Practice Exercise 10

1. The temperature, in ºC, on each day of November was recorded and the results summarized in a frequency table as shown below.

Temperature	Frequency
17	1
18	2
19	4
20	7
21	6
22	6
23	4

Draw:

(a) a frequency histogram

(b) a frequency polygon of the distribution.

2. Given below are the data about the height of 18 students.

Height	155-159	160-164	165-169	170-174	175-179
Frequency	4	7	10	5	2

(a) Find the midpoint or class centres of each class intervals.

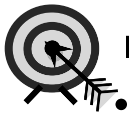
Height	**Class centre**	**Frequency**
155-159		4
160-164		7
165-169		10
170-174		5
175-179		2

(b) Draw a frequency histogram and polygon for the data given.

CORRECT YOUR WORK. ANSWERS ARE AT THE END OF TOPIC 2

Lesson 11: Cumulative Frequency Tables and Graphs

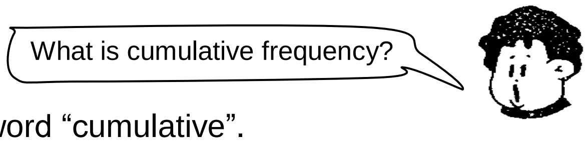
You learnt to identify and to draw histogram and frequency polygon in the previous lessons.



In this lesson, you will:

- define cumulative frequency
- identify a cumulative frequency table
- calculate cumulative frequencies
- define and present cumulative frequency graphs
- present information on a cumulative frequency graph.

In presenting data, sometimes we want to point out not the number of observations in a given class but the number falling below or above a specified value. A cumulative frequency distribution is then constructed.



First let us define the word “cumulative”.

Cumulative means “how much so far”. Think of the word **accumulate** which means **to gather together.**

To have cumulative totals, just **add up the values as you go.**

Example 1

Polo has earned this much in the last six month

Months	Earned
March	K120
April	K50
May	K110
June	K100
July	K50
August	K20

To work out the cumulative totals, just add up as you go.

The first line is easy, the total earned so far is the same as Polo earned that month.

Months	Earned
March	K120

But for April, the total earned so far is K120 + K50 = **K170**

Months	Earned	Cumulative
March	K120	K120
April	K50	**K170**

And for May we continue to add up: K170 + K110 = **K280**

Months	Earned	Cumulative
March	K120	K120
April	K50	K170
May	K110	**K280**

Notice how we add the previous month's cumulative total to this month's earnings?

Here is the calculation for the rest of the months.

- June is K280 + K100 = **K380**
- July is K380 + K50 = **K430**
- August is K430 + K20 **= K450**

And this is the result.

Months	Earned	Cumulative
March	K120	K120
April	K50	K170
May	K110	K280
June	K100	K380
July	K50	K430
August	K20	K450

The last cumulative total should match the total of all earnings.

K450 is the last cumulative total …

...it is also the total of all earnings.

K120 + K50 + K110 + K100 + K50 + K 20 = **K450**

So, we got it right.

So that's how to do it, add up as you go down the list and you will have the cumulative totals.

We can now define cumulative frequency.

Cumulative frequency is the total frequency up to a given data value.

The cumulative frequency of each score is found by adding the frequencies of all the scores up to and including that particular score.

You can think of a cumulative frequency as a **running total**.

For continuous data, cumulative frequency is the total frequency up to a given class boundary or class limit.

Example 2

The heights of 96 girls in Year 9 were recorded.

Height in cm	Frequency	Cumulative frequency
$120 \le h < 130$	1	1
$130 \le h < 140$	5	1 + 5 = 6
$140 \le h < 150$	18	6 + 18 = 24
$150 \le h < 160$	31	24 + 31 = 55
$160 \le h < 170$	24	55 + 24 = 79
$170 \le h < 180$	13	79 + 13 = 92
$180 \le h < 190$	4	92 + 4 = 96
	Total = 96	

The total frequency should be the same as the last cumulative frequency

We can present and show the cumulative frequency with a graph.

An Ogive (cumulative frequency graph) is a graph that represents the cumulative frequencies of the classes in a frequency distribution. It shows the data below or above a particular value.

The ogive is a cumulation of frequencies by class intervals arranged in table.

There are two types of Ogives, These are:

(a) The Less Than Ogive
(b) The Greater Than Ogive.

Steps for constructing a Less Than Ogive chart or Less Than Cumulative Frequency graph.

1. Draw and label the horizontal and vertical axes.
2. Take the cumulative frequencies along the y axis (vertical axis) and the upper class limits on the x axis (horizontal axis)
3. Plot the cumulative frequencies against each upper class limit.
4. Join the points with a smooth curve.

Steps for constructing a greater than or more than Ogive chart (more than Cumulative frequency graph):

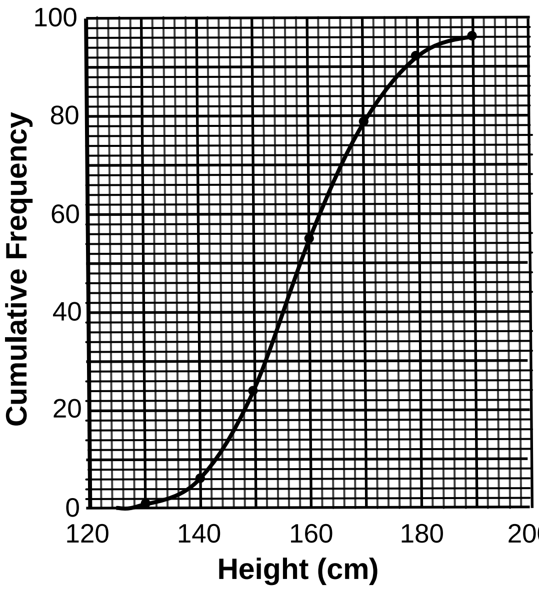
1. Draw and label the horizontal and vertical axes.
2. Take the cumulative frequencies along the y axis (vertical axis) and the lower class limits on the x axis (horizontal axis)
3. Plot the cumulative frequencies against each lower class limit.
4. Join the points with a smooth curve.

To draw a cumulative frequency graph you plot the cumulative frequencies against the corresponding class boundaries.

Look at the examples of “less than’ “and greater than” cumulative frequency curves of the heights of 96 girls in Year 9 as shown in Figure 1 and Figure 2 respectively.

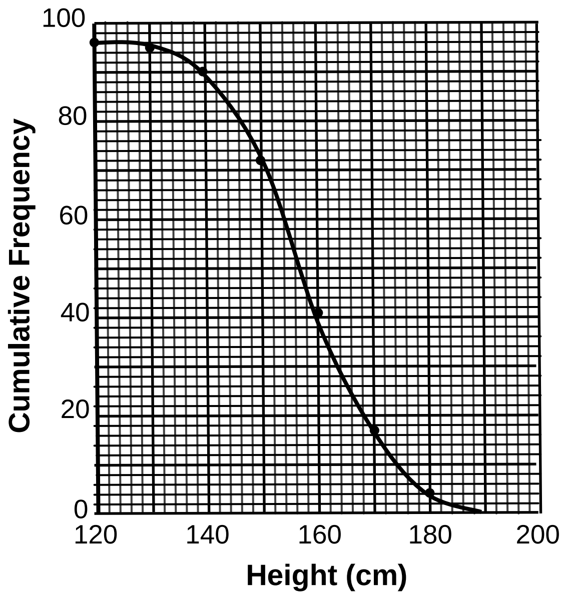
Height in cm	**Frequency**	**Less Than Cumulative Frequency (< Ogive)**	**Greater Than Cumulative Frequency (> Ogive)**
$120 \le h < 130$	1	1	96
$130 \le h < 140$	5	6	95
$140 \le h < 150$	18	24	90
$150 \le h < 160$	31	55	72
$160 \le h < 170$	24	79	41
$170 \le h < 180$	13	92	17
$180 \le h < 190$	4	96	4
	Total = 96		

Figure 1



THE LESS THAN (<) OGIVE

Figure 2



THE GREATER THAN (>) OGIVE

Cumulative frequency curves often have the distinctive **S** shape.

NOW DO PRACTICE EXERCISE 11

Practice Exercise 11

1. A medical practitioner at a Saint Mary's Clinic measured the heights of 100 patients. The measurements are recorded in the table.

Heights (cm)	Number of Patients	Class Limits		Cumulative frequency	
		Lower limits	Upper limits	Less than (<)	Greater Than (>)
168-170	4	168.5	170.5		
171-173	10				
174-176	14				
177-179	26				
180-182	22				
183-185	14				
186-188	7				
189-191	3				

(a) Determine the lower and upper limits. The first one is done for you.

(b) Calculate the cumulative frequencies.

(c) Draw the cumulative frequency curves.

2. Refer to the Frequency distribution table below.

Class interval	Frequency
25-29	3
30-34	6
35-39	8
40-44	14
45-49	19
50-54	17
55-59	13
60-64	9
65-69	7
70-74	4
	N = 100

(a) Expand the distribution table showing the class limits and cumulative frequencies.

Draw your table here.

Heights (cm)	Number of Patients	Class Limits		Cumulative Frequency	
		Lower limits	Upper limits	Less than (<)	Greater Than (>)
25-29	3				
30-34	6				
35-39	8				
40-44	14				
45-49	19				
50-54	17				
55-59	13				
60-64	9				
65-69	7				
70-74	4				
	N = 100				

(b) Draw the "Less than" and "Greater than" Ogive frequency curves.

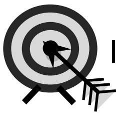
LESS THAN OGIVE FREQUENCY CURVE

GREATER THAN OGIVE FREQUENCY CURVE

CORRECT YOUR WORK. ANSWERS ARE AT THE END OF TOPIC 2

Lesson 12: Relative Frequency

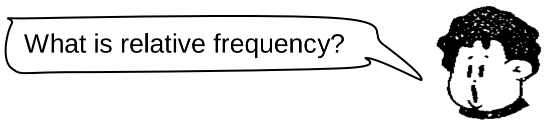
You learnt the meaning of cumulative frequency and cumulative frequency curves in the previous lessons. You also learnt to work out the cumulative frequencies and draw the ogives.



In this lesson, you will:

- define relative frequency
- compute relative frequencies
- fill in a relative frequency distribution table
- draw a relative frequency histogram

Sometimes the frequency distribution can be shown through the computation of the proportion of the frequency. This proportion of the frequency is called the relative frequency.



Relative frequency is defined as the measurement of data through a table showing the percentage in proportion of every frequency to the total frequency.

The relative frequency is calculated by using the formula:

$$\text{Relative Frequency (RF\%)} = \frac{\text{Frequency}}{\text{Total Frequency}}$$

For example:

Find the relative frequency from the given data below.

Class interval	Frequency
45-49	4
50-54	7
55-59	8
60-64	11
65-69	9
70-74	7
75-79	4
	N= 50

Solution:

Solve for the Relative Frequencies using the formula: $(RF\%) = \dfrac{\text{Frequency}}{\text{Total Frequency}}$

(a) For the class interval 45-49, $RF\ (\%) = \dfrac{4}{50}$

$= 0.08 \times 100$

= 8.00 or 8%

(b) For the class interval 50-54, $RF\ (\%) = \dfrac{7}{50}$

$= 0.14 \times 100$

= 14.00 or 14%

(c) For the class interval 55-59, $RF\ (\%) = \dfrac{8}{50}$

$= 0.16 \times 100$

= 16.00 or 16%

(d) For the class interval 60-64, $RF\ (\%) = \dfrac{11}{50}$

$= 0.22 \times 100$

= 22.00 or 22%

(e) For the class interval 65-69, $RF\ (\%) = \dfrac{9}{50}$

$= 0.18 \times 100$

= 18.00 or 18%

(f) For the class interval 70-74, $RF\ (\%) = \dfrac{7}{50}$

$= 0.14 \times 100$

= 14.00 or 14%

(g) For the class interval 75-79, $RF\ (\%) = \dfrac{4}{50}$

$= 0.08 \times 100$

= 8.00 or 8%

If the frequency of the frequency distribution table is changed into relative frequency then the frequency distribution table is called as **relative frequency distribution table.**

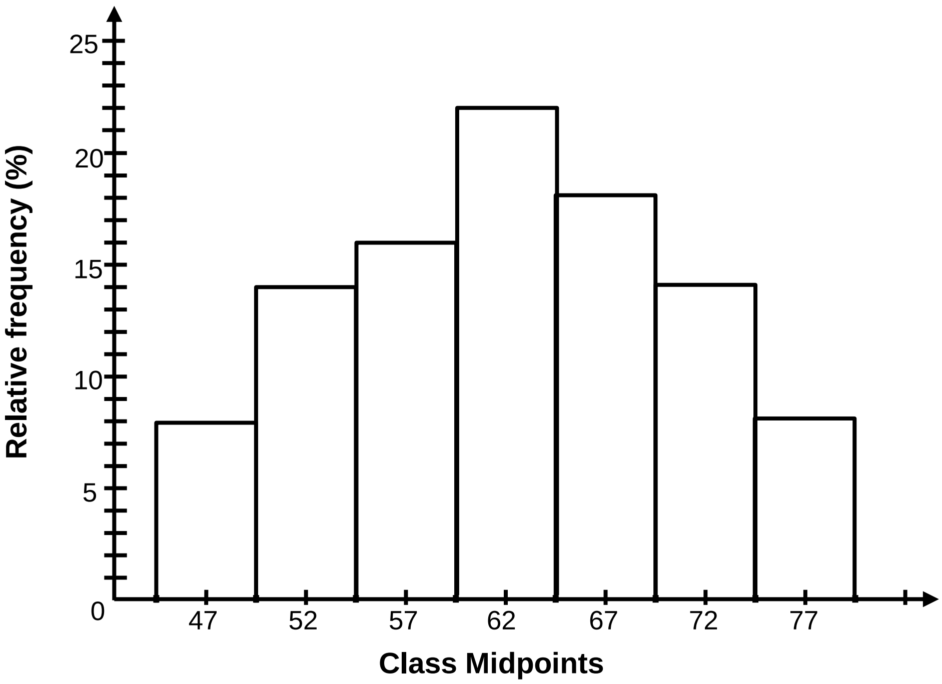
Here is the relative frequency distribution table of the given data.

Class interval	Frequency	RF (%)
45-49	4	8
50-54	7	14
55-59	8	16
60-64	11	22
65-69	9	18
70-74	7	14
75-79	4	8
	N = 50	**100**

We can also illustrate the relative frequency graph of the data above. It may be a histogram or a frequency polygon.

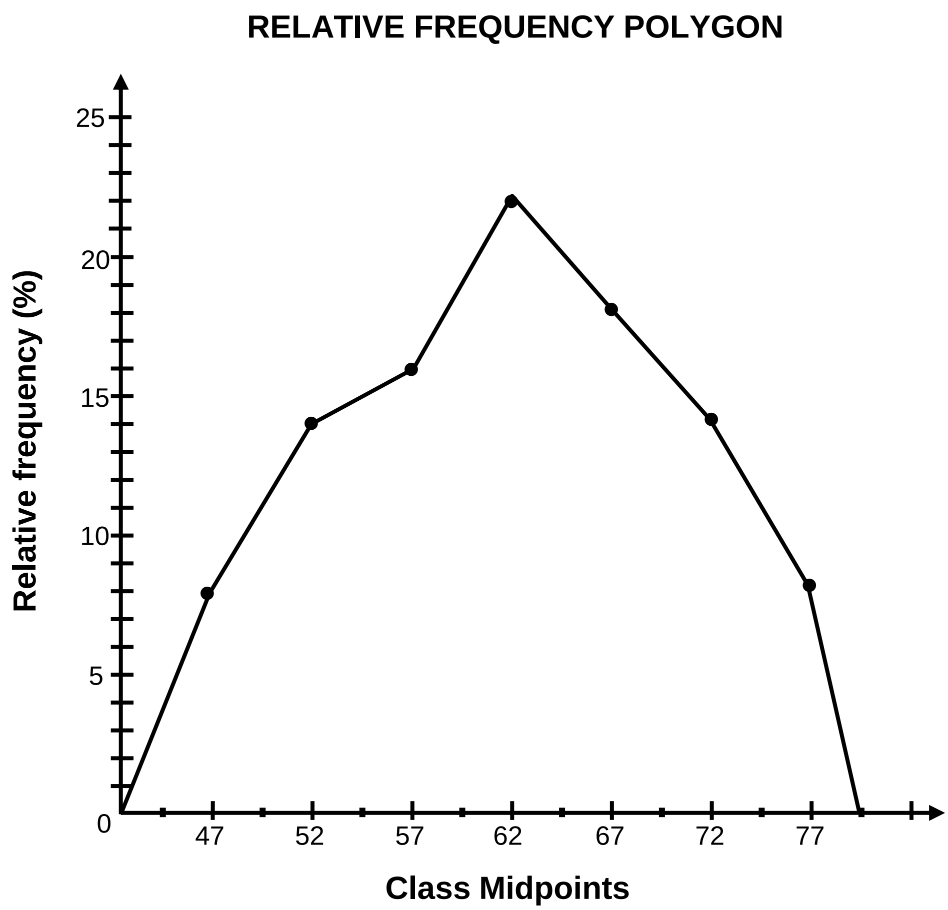
The diagram show the relative frequency histogram

RELATIVE FREQUENCY HISTOGRAM



Notice that the bars are always vertical and are placed next to each other without gaps. The relative frequencies are shown on the vertical axis and the class marks or midpoints on the horizontal axis.

The relative frequency polygon is constructed by plotting the class marks or midpoints with the relative frequencies and joining the points with a line. If a histogram had been drawn, just get the midpoints of the bars on top and connect the points with a line. The polygon closes by extending the endpoints of the line segments to the next class mark. See diagram on the next page.



NOW DO PRACTICE EXERCISE 12

Practice Exercise 12

1. Below is a frequency distribution of scores in a Mathematics examination.

Class interval	**Frequency**
15-19	1
20-24	3
25-29	6
30-34	10
35-39	13
40-44	23
45-49	15
50-54	12
55-59	12
60-64	5
	N = 100

(a) Identify the midpoints of each class intervals

Class interval	**Frequency**	**Midpoints**
15-19	1	
20-24	3	
25-29	6	
30-34	10	
35-39	13	
40-44	23	
45-49	15	
50-54	12	
55-59	12	
60-64	5	
	N = 100	

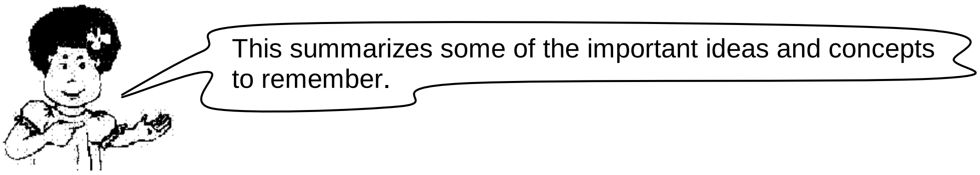
(b) Find the relative frequency for each class interval

Class interval	**Frequency**	**R% Frequency**
15-19	1	
20-24	3	
25-29	6	
30-34	10	
35-39	13	
40-44	23	
45-49	15	
50-54	12	
55-59	12	
60-64	5	
	N = 100	

(c) Draw the relative frequency histogram and polygon.

CORRECT YOUR WORK. ANSWERS ARE AT THE END OF TOPIC 2

TOPIC 2: SUMMARY



- A Pictograph is a graph which uses pictures or symbols to represent statistical data. To make a pictograph follow the steps below:
  1. List each category
  2. If necessary, round the data to nearest whole numbers
  3. Choose a picture or symbol that can represent the number in each category.
  4. Choose a key
  5. Draw pictures to represent the number in each category
  6. Label the pictograph. Write the title and the key.
- Bar graphs are graphs which use parallel bars with equal width to show statistical data. The length of the bars is drawn proportional to the quantities they represent. The bars are drawn vertically and horizontally. When the bars are drawn vertically, the bar graph is called a Column graph or vertical bar graph.
- A Compound graph is a special type of bar graph that compares two or more quantities simultaneously on a graph. When a compound graph is drawn with different bars **beside each other** it is called a **compound bar graph**. If the different bars are drawn **on top of each other**, the compound graph is called a **stacked bar graph.**
- A **Histogram** is a special type of bar graph where the bars are always vertical and are placed next to each other without gaps. It is a way of representing a frequency distribution of data where the values of the variables or scores are placed on the horizontal axis and the frequency of the scores on the vertical axis.
- A **Frequency Polygon** is a special type of line graph which is drawn by joining all the midpoints of the top of the bars of a histogram. The area of the frequency polygon is equal to the area of the frequency histogram.
- A **Cumulative Frequency** is the total frequency up to a given data value.
- **An Ogive or Cumulative Frequency Graph** is a graph that represents the cumulative frequencies of the classes in a frequency distribution. It shows the data below or above a particular value.

  There are two types of Ogives:

  1) **The Less than Ogive** showing the data below a particular value
  2) **The Greater than Ogive** which shows the data above a particular value.
- **Relative Frequency** is the measurement of the data through a table showing the percentage in proportion of every frequency to the total frequency.

$$\boxed{\text{RF\%} = \frac{\text{Frequency}}{\text{Total Frequency}}}$$

REVISE LESSONS 13-18. THEN DO TOPIC TEST 3 IN ASSIGNMENT BOOK 3.

ANSWERS TO PRACTICE EXERCISES 7-12

Practice Exercise 7

1. (a) 300, 400, 250, 200, 500, 450, 350
   (b) Thursday
   (c) 2450
   (d) Wednesday
   (e) 300

2. (a) 450
   (b) 350
   (c) Chocolate Peanut Butter
   (d) Yes

3. (a) 12
   (b)

SHAWN'S FRIENDS HOBBY	
Computers	☺ ☺ ☺ ☺
Football	☺ ☺ ☺ ☺ ☺ ☺
Music	☺ ☺
Others	☺ ☺ ☺

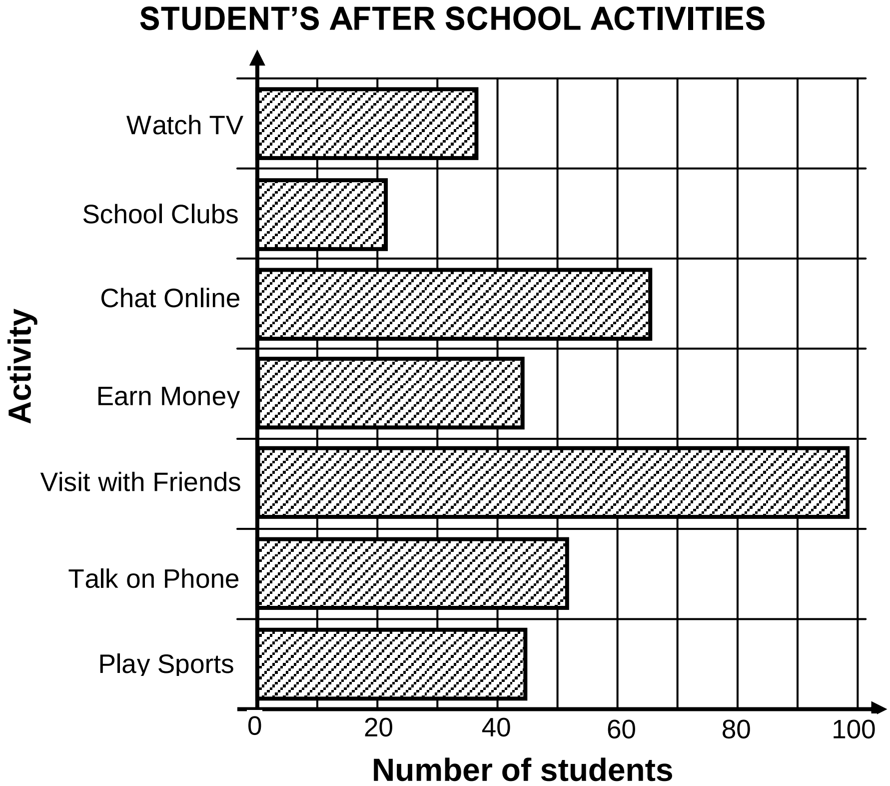
KEY: ☺ = 3 persons

Practice Exercise 8

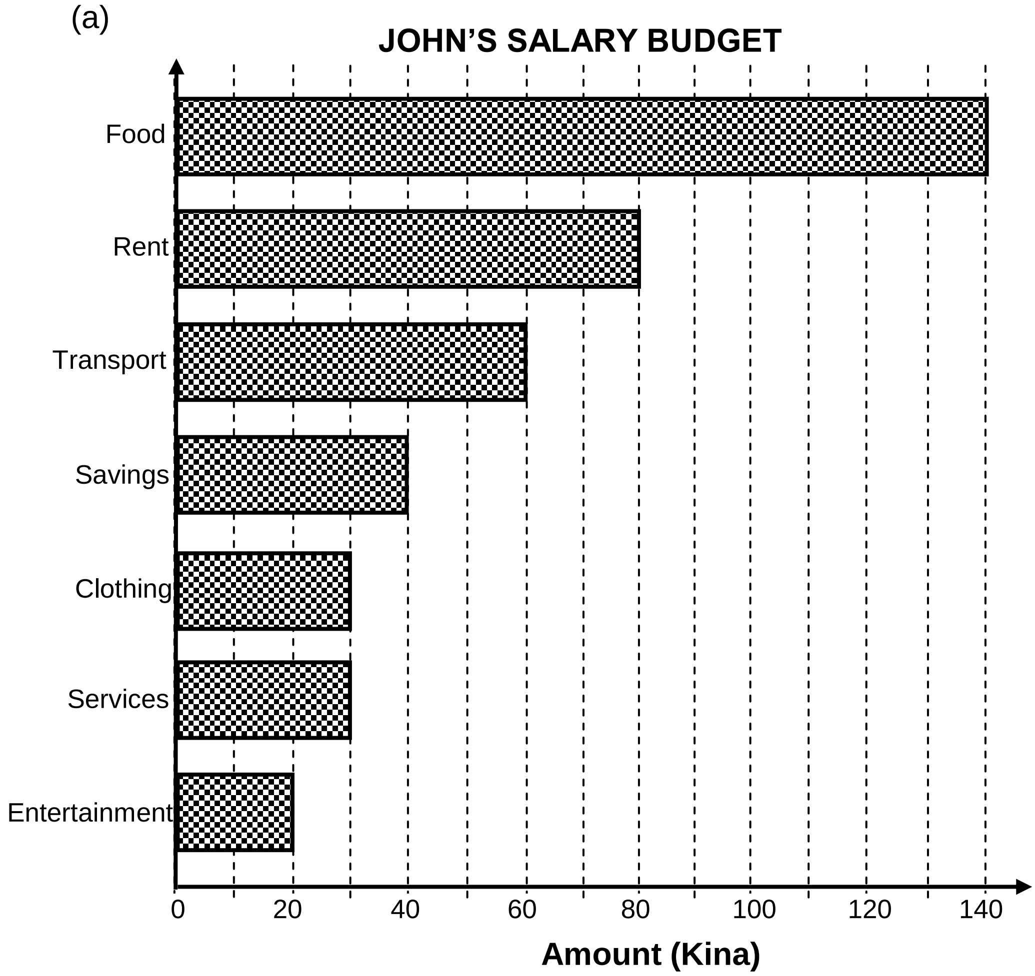
1. (a) 3 000 000
   (b) 800 000
   (c) Yes
   (d) migration

2. (a) i. Visit with Friends
   ii. School Clubs
   iii. 53
   i. 44
   ii. Visit with Friends, Chat Online, Talk on Phone, Play sports, Earn Money, Watch TV, School Clubs

(b)



3. (a)



(b) i. Food ii. Entertainment iii. 20%

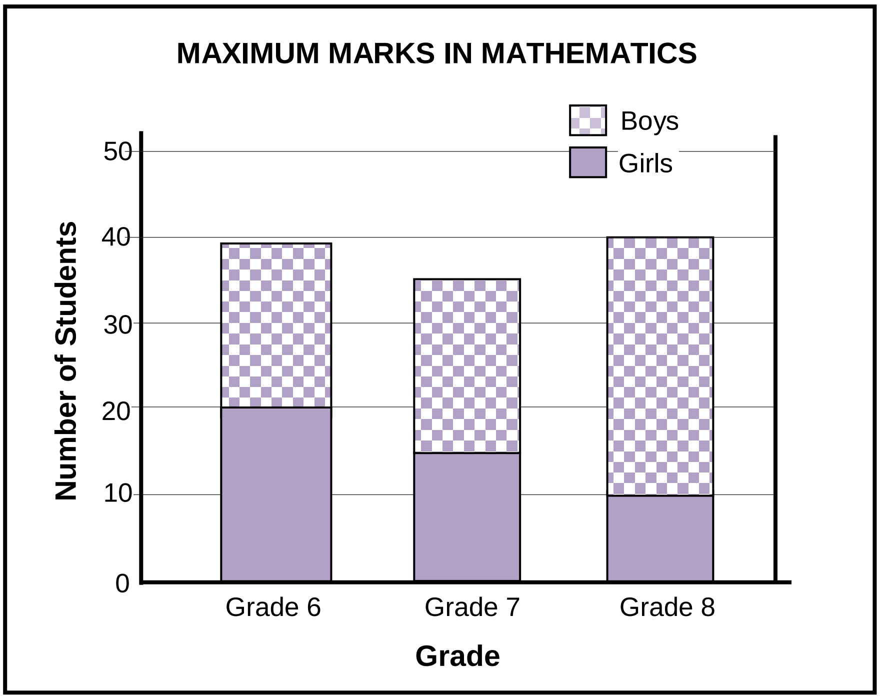
Practice Exercise 9

1. (a) Housewife
   (b) Awin
   (c) more
   (d) farmer, housewife and unskilled workers
   (e) 50
   (f) i. true ii. True

2. (a) White bars
   (b) 4 years
   (c) 2006-2007
   (d) 2010

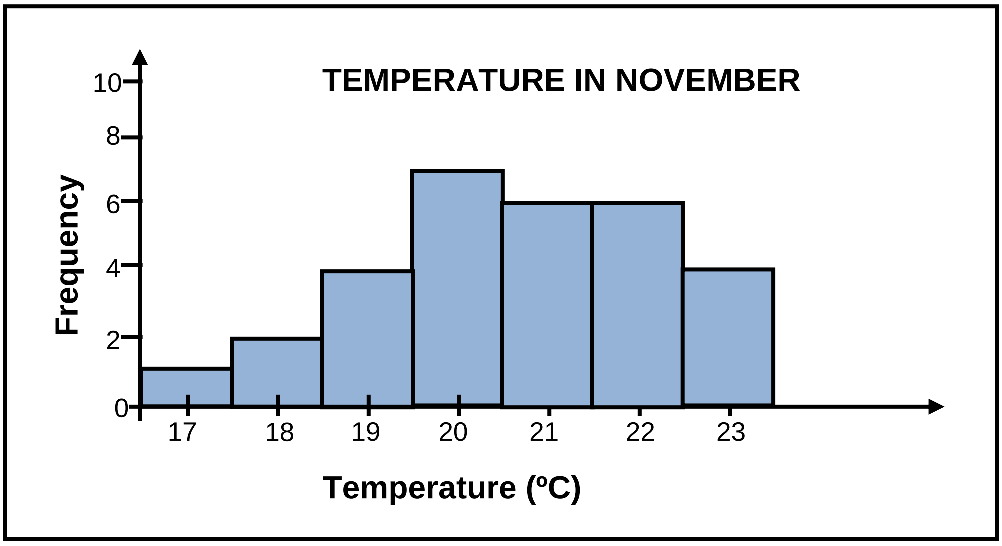
3. (a) Machinery
   (b) 570 to 600 million kina
   (c) (i) Chemical imports increased between 1980 and 1983
   (d) Increased

4.

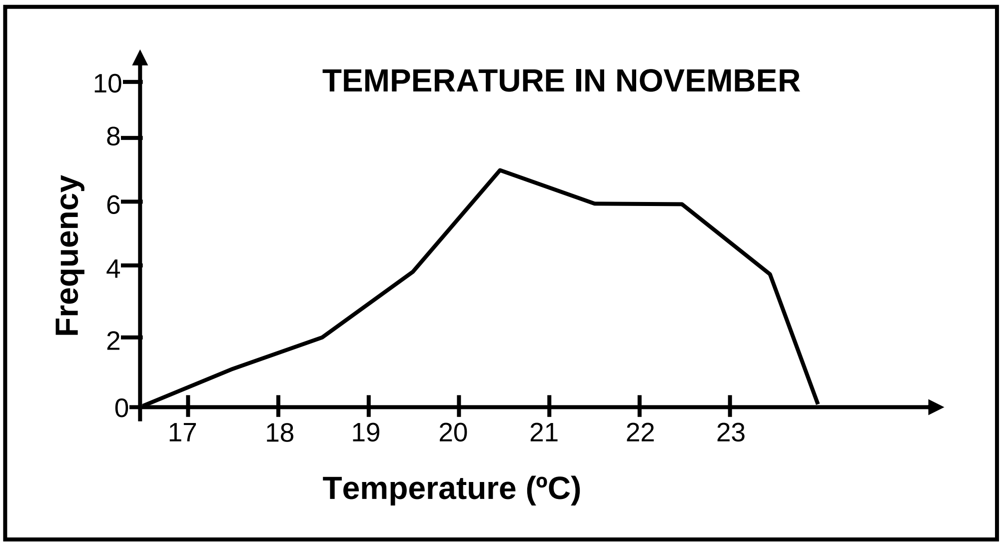


Practice Exercise 10

1. (a) Frequency histogram



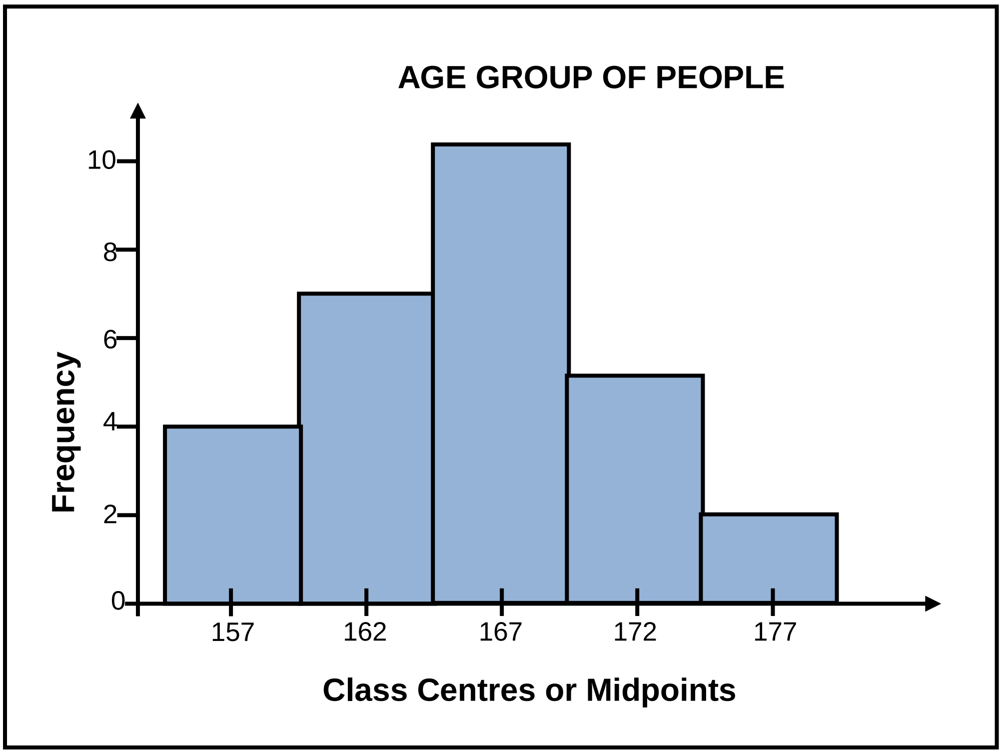
(b) Frequency Polygon



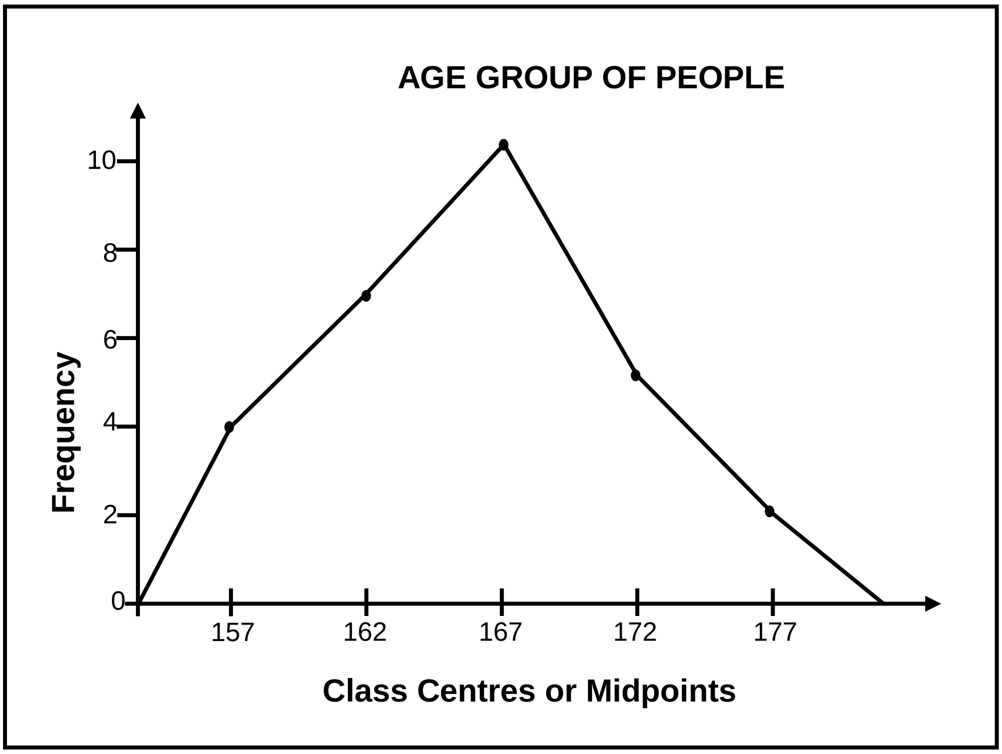
2. (a)

Height	Class centre	Frequency
155-159	**157**	4
160-164	**162**	7
165-169	**167**	10
170-174	**172**	5
175-179	**177**	2

(b) Frequency Histogram and Polygon



Histogram



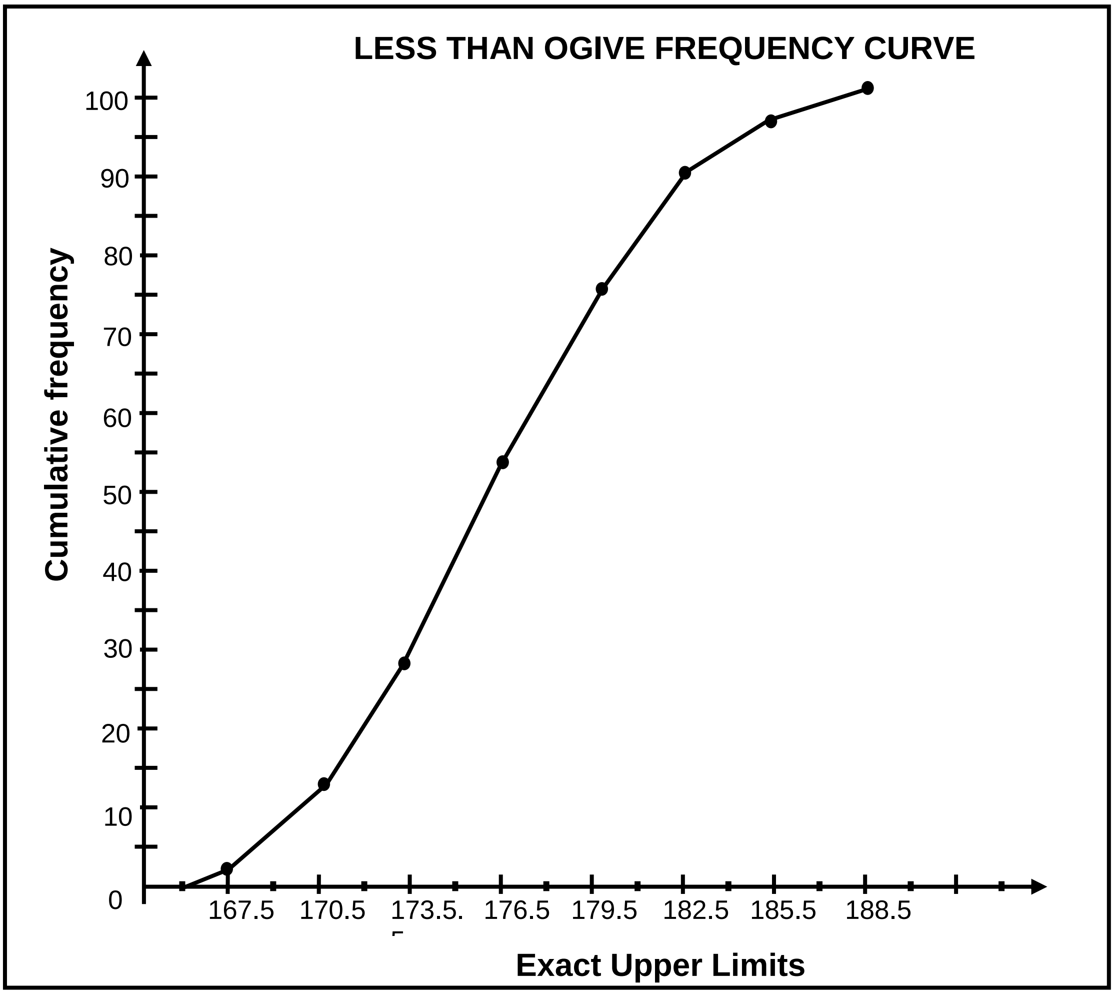
Polygon

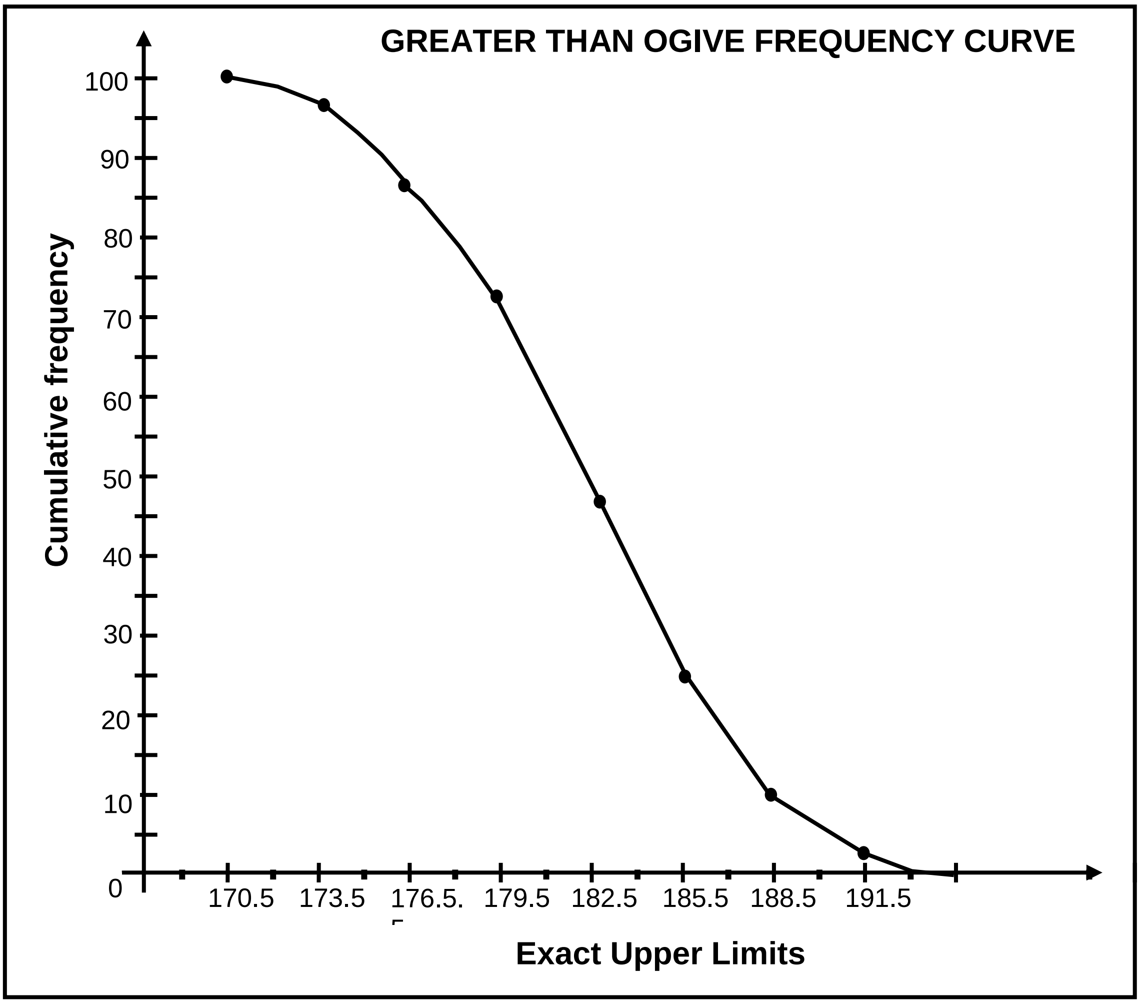
Practice Exercise 11

1. (a) (b)

Heights (cm)	Number of Patients	Class Limits		Cumulative frequency	
		Lower limits	Upper limits	Less than (<)	Greater Than (>)
168-170	4	167.5	170.5	**4**	**100**
171-173	10	**170.5**	**173.5**	**14**	**96**
174-176	14	**173.5**	**176.5**	**28**	**86**
177-179	26	**176.5**	**179.5**	**54**	**72**
180-182	22	**179.5**	**182.5**	**76**	**46**
183-185	14	**182.5**	**185.5**	**90**	**24**
186-188	7	**185.5**	**188.5**	**97**	**10**
189-191	3	**188.5**	**191.5**	**100**	**3**

(c)

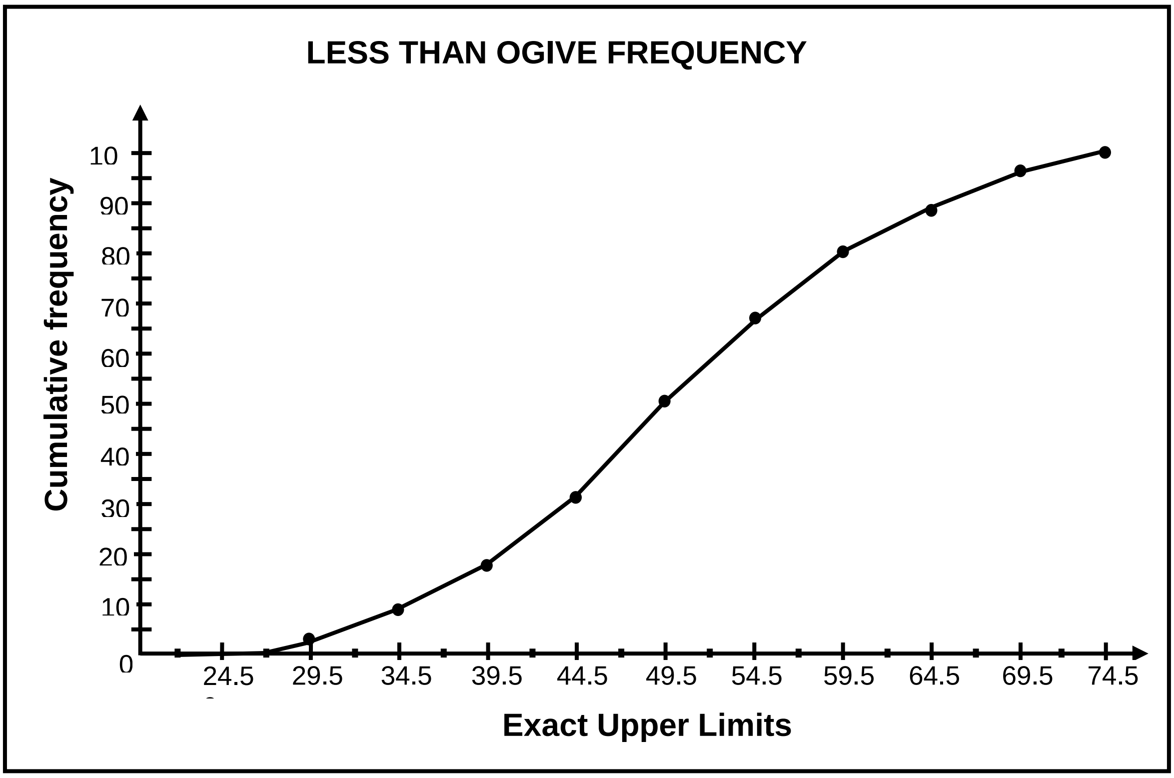


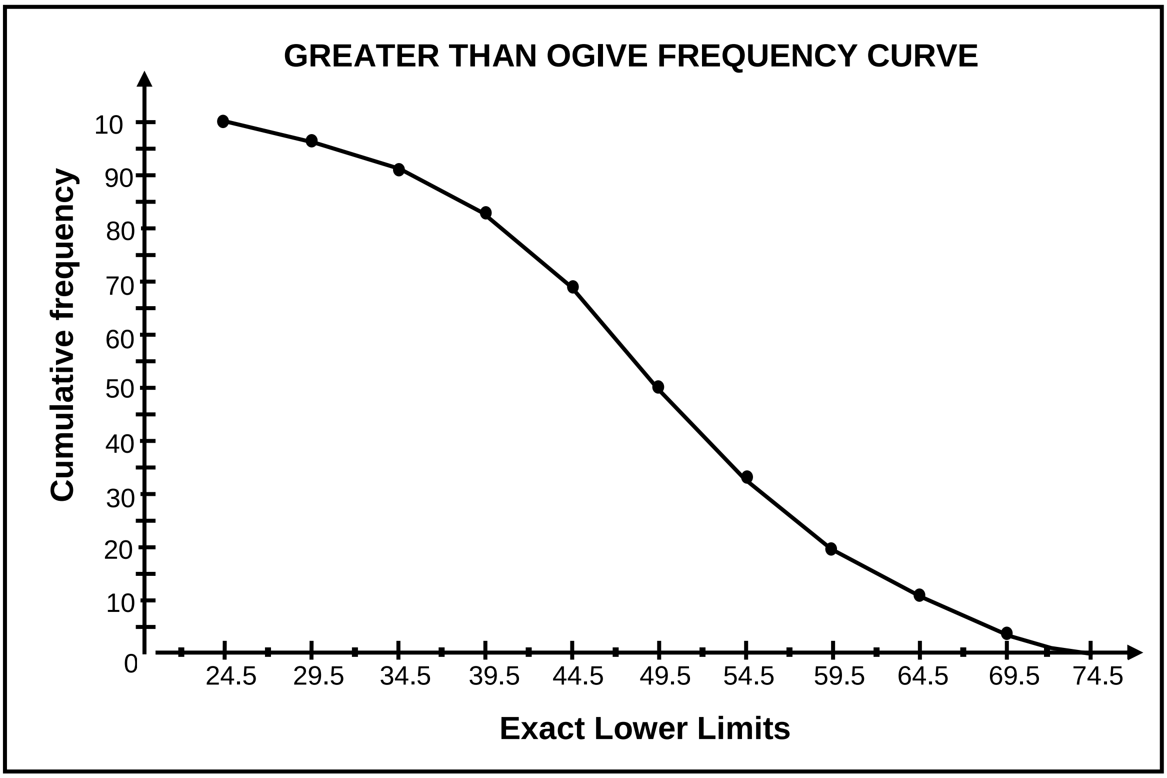


.2. (a)

Heights (cm)	Number of Patients	Class Limits		Cumulative Frequency	
		Lower limits	Upper limits	Less than (<)	Greater Than (>)
25-29	3	24.5	29.5	3	100
30-34	6	29.5	34.5	9	97
35-39	8	34.5	39.5	17	91
40-44	14	39.5	44.5	31	83
45-49	19	44.5	49.5	50	69
50-54	17	49.5	54.5	67	50
55-59	13	54.5	59.5	80	33
60-64	9	59.5	64.5	89	20
65-69	7	64.5	69.5	96	11
70-74	4	69.5	74.5	100	4
	N = 100				

(b) Draw the “Less than” and “Greater than” Ogive frequency curves.





Practice Exercise 12

1. (a) Midpoints

TABLES OF SCORES IN A MATHEMATICS TEST

Class interval	Frequency	Midpoints
15-19	1	**17**
20-24	3	**22**
25-29	6	**27**
30-34	10	**32**
35-39	13	**37**
40-44	23	**42**
45-49	15	**47**
50-54	12	**52**
55-59	12	**57**
60-64	5	**62**
	N = 100	

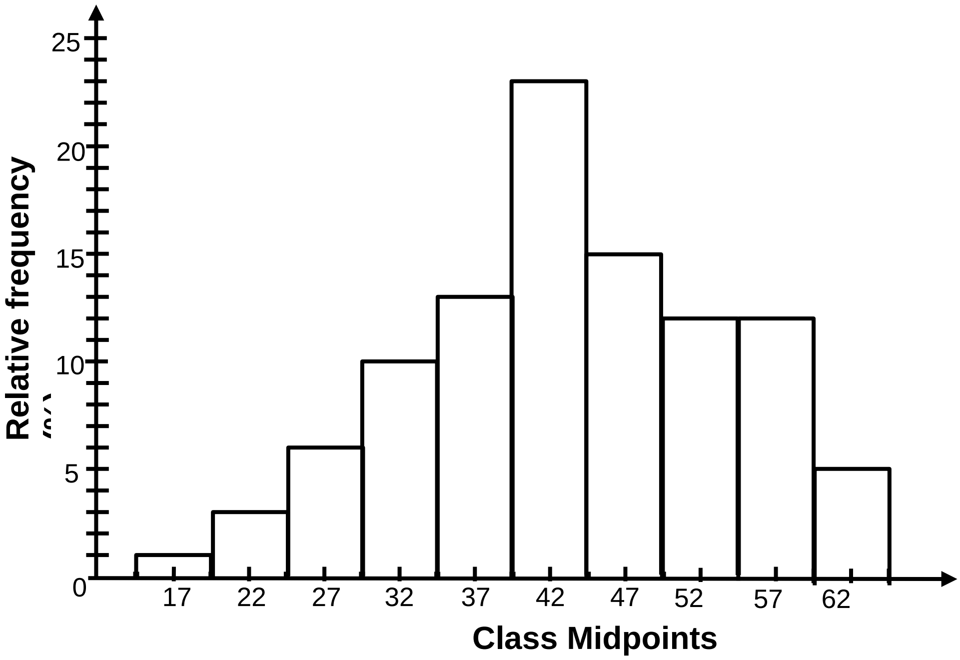
(b) Relative frequency for each class interval.

TABLES OF SCORES IN A MATHEMATICS TEST

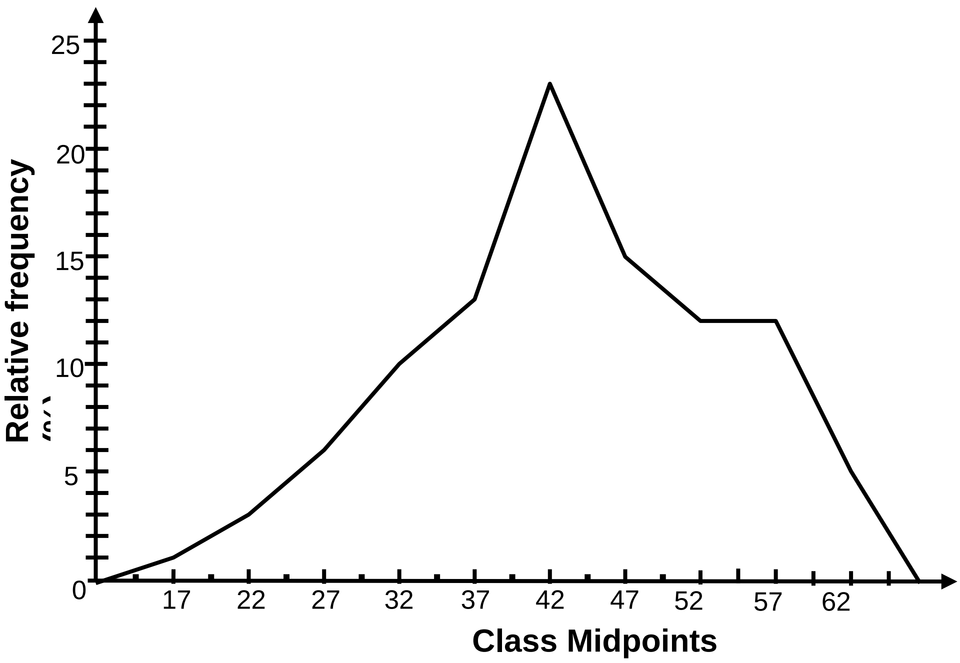
Class interval	Frequency	R% Frequency
15-19	1	**0.01 = 1%**
20-24	3	**0.03= 3%**
25-29	6	**0.06 = 6%**
30-34	10	**0.1 = 10%**
35-39	13	**0.13 = 13%**
40-44	23	**0.23 = 23%**
45-49	15	**0.15 = 15%**
50-54	12	**0.12 = 12%**
55-59	12	**0.12 = 12%**
60-64	5	**0.05 = 5%**
	N = 100	

(c)

RELATIVE FREQUENCY HISTOGRAM



RELATIVE FREQUENCY POLYGON



END OF TOPIC 2

TOPIC 3

MEASURES OF CENTRAL TENDENCY

TOPIC 3: MEASURES OF CENTRAL TENDENCY

Introduction

A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within the set of data. Sometimes, the measure of central tendency is called the measure of central location. They are referred to as "summary statistics".

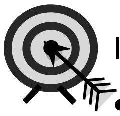
The mean (often called the average) is most likely the measure of central tendency that you are most familiar with but there are others such as the median and the mode.

There are three main measures of central tendency. These are the mean, the median and the mode. These three are all the valid measures of central tendency, but under different conditions, some measures of central tendency become more appropriate to use than others.

In this topic, we will look at the mean, median and mode and how to calculate them and consider the conditions under which they are most appropriate to be used.

Lesson 13: Mean of Ungrouped Data

Welcome to Lesson 13 of Unit 3. In the previous lessons, you learnt about the different types of data and how to work on them.



In this lesson, you will:

- define mean of ungrouped data
- use the formula to work out the mean of ungrouped data

You learnt that the mean of a set of numbers is often called the average in your Grade 7 and 8 Mathematics. You are now going to extend further your knowledge about the mean.

First, let us define ungrouped data.

> The data presented in its original form is known as **ungrouped data**. Ungrouped data is nothing but **raw data.**

The Mean of Ungrouped Data

The mean as you have learnt is another term for arithmetic average. It is the most popular and well known measure of central tendency. It can be used with both discrete and continuous data, although its used is most often with continuous data. If you have computed an average, you have computed a mean.

> **The mean or average of ungrouped data is simply the sum of all the values in the set of data divided by the number of values in the set of data.**

So when you have **N** values in a set of data and they have values $\mathbf{X_1}$, $\mathbf{X_2}$, ...$\mathbf{X_n}$, the mean denoted by $\mathbf{\overline{X}}$ (pronounced X bar) is:

$$\overline{X} = \frac{\mathbf{X_1 + X_2 + X_3 + \ldots X_6}}{\mathbf{N}}$$

Example 1

Suppose you have six scores:

12, 10, 18, 16, 20 and 14

If you let $X_1 = 12$; $X_2 = 10$; $X_3 = 18$; $X_4 = 16$; $X_5 = 20$ and $X_6 = 14$, the mean as represented by $\overline{X}$, is:

$$\overline{X} = \frac{X_1 + X_2 + X_3 + X_4 + X_5 + X_6}{N}$$

$$\overline{X} = \frac{12 + 10 + 18 + 16 + 20 + 14}{6}$$

$$\overline{X} = \frac{90}{6}$$

$\overline{X} = 15$ **Answer**

Instead of writing the equation for the mean as shown above, the equation is simplified in different manner using the Greek capital letter, ∑, pronounced "sigma" which means "summation or sum of".

$$\boxed{\overline{X} = \frac{\Sigma X}{N}}$$

where: $\overline{X}$ = the mean

ΣX = the sum of all the scores

N = the total number of scores

Sometimes you will calculate the mean for a set of numbers where many of the numbers are repeated. The shortcut explained below could save your time.

Example 2

Calculate the mean of these eight scores.

80 80 80 90 90 90 90 90

Solution:

To compute the mean, you could add the eight scores and then divide by 8

$$\overline{X} = \frac{X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8}{N}$$

$$\overline{X} = \frac{80 + 80 + 80 + 90 + 90 + 90 + 90 + 90}{8}$$

$$\overline{X} = \frac{690}{8}$$

$\overline{X} = 86.25$ **Answer**

Or, you could use this shortcut: $\boxed{\overline{X} = \frac{\Sigma fX}{N}}$

where: $\overline{X}$ = the mean

ΣfX = the sum of all the products of each score by the number of frequency

N = total number of scores

STEPS

1. Multiply each score by the number of times it occurs (f)(X)

$$80 \times 3 = 240, \ 90 \times 5 = 450$$

2. Add these products (ΣfX) 240 + 450 = 690

3. Compute the mean ($\overline{X}$).

$$\overline{X} = \frac{\Sigma fX}{N}$$

$$\overline{X} = \frac{690}{8}$$

$\overline{X}$ = 86.25 **Answer**

Example 3

Calculate the mean of the following heights in centimetres of 20 Boys.

165	180	160	173
180	168	175	170
162	170	170	162
178	165	170	165
173	178	168	173

To calculate the mean, the set of data may be presented on a frequency distribution table such as the one below, where each height is paired in the table with the number of times (the frequency) it occurred.

Heights (X)	**Frequency (f)**
160	1
162	2
165	3
168	7
170	4
173	3
175	1
178	2
180	2
	N = 20

We may expand the table to include a column for the product of the height and the frequency (f)(X). We add all these products and divide by the sum N.

Heights (X)	Frequency (f)	(f)(X)
160	1	160
162	2	324
165	3	495
168	7	336
170	4	680
173	3	519
175	1	175
178	2	356
180	2	360
	N = 20	ΣfX = 3405

Now you can calculate the mean.

Solution:

$$\overline{X} = \frac{\Sigma fX}{N}$$

$$\overline{X} = \frac{3405}{20}$$

$\overline{X}$ = 170.25 cm **Answer**

NOW DO PRACTICE EXERCISE 13

Practice Exercise 13

1. Find the mean of the following set of data.

   a) 75 95 100 85 80

   b) 21 26 25 21 28 27

   c) 47 70 60 70 105

2. On a four day trip, Lucy's family drove 240, 100, 200 and 160 kilometres.

   What is the mean number of kilometres they drove for day?

3. Jason received these scores on Math tests:

   85 70 80 90 80 80 80 75 85 75 90.

   Find Jason's mean score.

4. Julio is on the track team. He recorded the kilometres he ran each day for the past week as follows:

5.9 km , 6 km, 3.7 km, 4.5 km, 6.2 km, 6.1 km and 3.8 km

To the nearest tenth of a kilometre, what was the mean number of kilometres he ran a day?

---

5. The list below shows the number of rainy days in a certain province in 2008.

January	10	July	14
February	9	August	18
March	12	September	13
April	8	October	11
May	12	November	8
June	15	December	9

a) Arrange the scores from highest to lowest.

b) Complete the frequency table by filling in the columns.

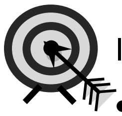
Days (X)	Frequency (f)	(f)(X)
18		
15		
14		
13		
12		
11		
10		
9		
8		
	N =	ΣfX =

d) Calculate the mean.

CORRECT YOUR WORK. ANSWERS ARE AT THE END OF TOPIC 3.

Lesson 14: Mean of Grouped Data

In Lesson 13, you learnt to calculate the mean of ungrouped data.



In this lesson, you will:

- define mean of grouped data
- use the formula to work out the mean of ungrouped data

Sometimes the data for which we want to find a mean has been grouped into classes. We don't know the individual values, only the number of values in each class.

Data which have been arranged in groups or classes rather than showing all the original figures are called grouped data.

When you are given data which has been grouped, you can't work out the mean exactly because you don't know what the values are exactly (you just know that they are between certain values).

However, we calculate an estimate of the mean with the formula:

$$\boxed{\overline{X} = \frac{\Sigma fM}{N}}$$

f = the frequency

where:

M = the midpoint of the group

fM = the product of the frequency and each midpoint

∑ = means 'the sum of'

N = total number of scores

Example 1

Suppose you have the list of ratings of 50 students in a Statistics Class in a certain school as shown in Table 14.1 on the next page. Grades are quantified by making an A equal to 6,
B equal to 5, C equal to 4 D equal to 3, and E equal to 2 as shown in the first column.

Your task is to find the average of the grades of the students. The values in the fourth column are the products of the row values in each of second and third column (fM); that is 6 x 10 = 60, 5 x 13 = 65, 4 x 12 = 48, 3 x 5 = 15 and 2 x 10 = 20.

Then the fourth column is summed up to apply the formula : $\overline{X} = \frac{\Sigma fM}{N}$

Table 14.1

Class Interval	Midpoints(M)	Frequency(f)	Fm
A 6	6	10	60
B 5	5	13	65
C 4	4	12	48
D 3	3	5	15
E 2	2	10	20
		N = 50	ΣfM = 208

Solution:

$$\overline{X} = \frac{\Sigma fM}{N}$$

$$\overline{X} = \frac{208}{50}$$

$$\overline{X} = 4.2 \longleftarrow \textbf{Answer to one dec. place}$$

The above computation may be used when the class interval is equal to 1.

When the interval size is greater than 1, the method used is given on the following examples.

Example 2

Consider the times taken by 30 students to do a test. Their times have been summarised in Table 14.2 below.

Table 14.2

TIME TAKEN BY 30 STUDENTS

Minutes spent on test	Number of students (the Frequency, f)
0 to less than 5 minutes	2
at least 5 but less than 10 minutes	12
at least 10 but less than 20 minutes	16

We make the assumption that within each class the mean of the values in that class equals the mid-point value of the class.

To find the mid-point value for each class add the values of the 2 end points together and divide by 2. The formula is:

$$\boxed{M = \frac{LS + HS}{2}}$$

where: M = midpoint

LS = the lowest score in the class interval

HS = the highest score in the class interval

Example

a) $\dfrac{0+5}{2} = \dfrac{5}{2} = 2.5$ b) $\dfrac{5+10}{2} = \dfrac{15}{2} = 7.5$ c) $\dfrac{10+20}{2} = \dfrac{30}{2} = 15$

We may expand the preceding frequency table to include a column for the midpoints and the number of students and the midpoints (fM) as shown in Table 14.2.1. We add all these products to get ΣfM.

Table 14.2.1

Minutes spent on test	Number of students (the Frequency, f)	Midpoint (M)	fM
0 to less than 5 min	2	2.5	5
At least 5 but less than 10 min	12	7.5	90
At least 10 but less than 20 min	16	15	240
	N = 30		ΣfM = 335

To find the mean, apply the formula by substituting the values of ΣfM and N.

Solution:

$$\overline{X} = \frac{\Sigma fM}{N}$$

$$\overline{X} = \frac{335}{30}$$

$\overline{X} = 11.2$ ⟵ **Answer to one dec. place**

Example 3

Find the mean of the Scores in a Revision Test of 42 students shown on the frequency table below.

Table 14.3

SCORES IN A REVISION TEST

Class Interval	Frequency(f)
61-65	2
66-70	3
71-75	7
76-80	9
81-85	10
86-90	6
91-95	4
96-100	1
	N = 42

First, we have to determine the midpoints or the middle score of each class intervals. As you have learnt, the midpoint is computed by the formula:

$$\boxed{M = \frac{LS + HS}{2}}$$

where: M = midpoint

LS = the lowest score in the class interval

HS = the highest score in the class interval

Illustrative example for the first class interval:

$$M = \frac{LS + HS}{2}$$

$$M = \frac{61 + 65}{2}$$

$$= \frac{126}{2}$$

$$= 63$$

The third column of Table 14.4 shows all the midpoints or the middle scores of each class interval (M).

Table 14.4

Class Interval	Frequency (f)	Midpoints (M)
61-65	2	63
66-70	3	68
71-75	7	73
76-80	9	78
81-85	10	83
86-90	6	88
91-95	4	93
96-100	1	98
	N = 42	

Now, get the product of each midpoint and the corresponding frequency within its interval (fM).

Illustrative examples

For the first class interval 61-65 we have: $fM_1 = 2 \times 63 = 126$;

For the interval 66-70 we have; $fM_2 = 3 \times 8 = 204$; the third, $fM_3 = 7 \times 73 = 511$; the fourth, $fM_4 = 9 \times 78 = 702$ and so on.

The fourth column on Table 14.5 shows the products of the midpoint for each class intervals and the corresponding frequency (fM).

Column 2 and 4 are summed up to get N and ΣfM.

Table 14.5

Class Interval	Frequency (f)	Midpoints (M)	fM
61-65	2	63	126
66-70	3	68	204
71-75	7	73	511
76-80	9	78	702
81-85	10	83	830
86-90	6	88	528
91-95	4	93	372
96-100	1	98	98
	N = 42		ΣfM= 3371

Substituting the values of N and ΣfM in the formula, we can now calculate the mean score.

Solution: $\overline{X} = \dfrac{\Sigma fM}{N}$

$$\overline{X} = \frac{3371}{42}$$

$\overline{X} = 80.3$ ← **Answer to one dec. place**

The foregoing computation has been made easy following the steps below.

1. Determine the midpoints of each class interval.
2. Get the product of each midpoint and the corresponding frequency within its interval to obtain ΣfM.
3. Apply the formula by substituting the values ΣfM and N.

NOW DO PRACTICE EXERCISE 14

Practice Exercise 14

1 Refer to the weight in kilograms of students in a certain class listed below.

39	45	44	41	40	46
48	42	42	45	42	40
39	43	49	49	49	42

a) Construct a grouped frequency table for these data consisting of 6 class intervals.

Class Interval	Frequency (f)	Midpoints (M)	fM
39-40			
41-42			
43-44			
45-46			
47-48			
49-50			
	N		ΣfM

b) Find the Mean.

2. Refer to the following data.

10	16	5	3	11	8
9	15	12	14	16	18
20	20	18	16	19	14
14	17	13	10	16	10
7	10	12	6	8	5

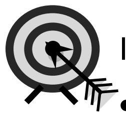
a) Construct a frequency distribution table.

b) Find the mean.

CORRECT YOUR WORK. ANSWERS ARE AT THE END OF TOPIC 3

Lesson 15: Median of Ungrouped Data

You learnt to calculate the mean of grouped data in the last lesson..



In this lesson, you will:

- define median of ungrouped data
- use the formula to work out the median of ungrouped data

As we have already learnt, the data presented in its original form is known as ungrouped data. Ungrouped data is nothing but raw data.

The **Median** is defined as the centre value in an ordered set of data in the distribution. It is the point in the distribution below which 50% of the scores lie.

.
That is, the **median of a distribution is the value which divides it into two equal parts**. It is the value such that the number of observations above it is equal to the number of observations below it. In finding the median, therefore, **the data must be arranged** in **ascending or descending order** of magnitude. The median is the point on a score scale that is the middle area of the histogram. One-half of the area of the histogram will fall below the median and one-half will fall above it.

Median of Ungrouped Data

When the set of data (**n**) is odd in number, the median is the $\left[\frac{n+1}{2}\right]^{th}$ score counted either from the top or from the bottom of the distribution.

For example, if n is 19, the median is the 10^{th} score, counted from the highest or from the lowest.

Thus, the formula in finding the median of ungrouped data if **n** is odd is:

$$\boxed{\textbf{Median} = \mathbf{X}_{\left(\frac{n+1}{2}\right)}}$$

When the set of data (n) is even, the median is the average between the $\left(\frac{n}{2}\right)^{th}$ score and the $\left(\frac{n}{2}+1\right)^{th}$ score. In other words it is the mean of the two middle values. This places the median in the middle of these two values. So, if n = 6, the median is the average of the third and the fourth scores.

For example, the set of data: 3, 5, 7, 10, 12, and 13 will have a median which is midway from 7 and 10 which is 8.5.

Since $\frac{6}{2}$ = 3, then the $\left(\frac{n}{2}\right)^{th}$ score is 7 and since 3 + 1= 4, then the $\left(\frac{n}{2}+1\right)^{th}$ score is 10.

This means that the median lies between the third and the fourth scores

The formula for finding the median of ungrouped data, if n is even is:

$$\boxed{\textbf{Median} = \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n+2}{2}\right)}}{2}}$$

Now, study the following examples.

Example 1

Find the median of the following set of scores.

23, 24, 25, 25, 26, 27, 28, 28, 30

Solution: Since n = 9 and it is odd, we use the formula:

$$\text{Median} = X_{\left(\frac{n+1}{2}\right)}$$

$$\text{Median} = X_{\left(\frac{9+1}{2}\right)}$$

$$= X_{\left(\frac{10}{2}\right)}$$

$$= X_5$$

This means that the median is the fifth score.

Therefore, the median is the fifth score which is **26.**

Example 2

Find the median of the following set of scores.

3, 8, 9, 11, 12, 18, 22, 31

Solution: Since n = 8 and it is even, we use the formula;

$$\text{Median} = \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n+2}{2}\right)}}{2}$$

$$\text{Median} = \frac{X_{\left(\frac{8}{2}\right)} + X_{\left(\frac{8+2}{2}\right)}}{2}$$

$$= \frac{X_4 + X_5}{2}$$

This means that the median is between the 4^{th} and 5^{th} values.

As we have learned, if n is even, the median is the mean or average of the two middle scores.

So, if we find the average of the two scores which are 11 and 12, we have

$$\text{Median} = \frac{11 + 12}{2}$$

$$= \frac{23}{2}$$

$$= 11.5$$

Therefore, the median of the set of scores is **11.5.**

Example 3

Find the median of the following raw scores.

12, 15, 19, 21, 6, 4, 2

Solution: First, arranged the scores in ascending or descending order.

2, 4, 6, 12, 15, 19, 21

Since n = 7 which is odd, we use the formula:

$$\text{Median} = X_{\left(\frac{n+1}{2}\right)}$$

$$\text{Median} = X_{\left(\frac{7+1}{2}\right)}$$

$$= X_{\left(\frac{8}{2}\right)}$$

$$= X_4$$

This means that the median is the fourth score.

Therefore, the median is **12.**

NOW DO PRACTICE EXERCISE 15

Practice Exercise 15

1. Find the median of the following sets of data.

   a. 31, 36, 35, 38, 33, 37, 32

   b. 75, 72, 77, 73, 79, 76, 75, 72

   c. 106, 102, 111, 105, 109, 103

2. Refer to the following set of data.

   256, 343, 219, 251, 121, 283, 346, 291, 462, 169, 201, 232, 198, 305

   a) Arrange the data in ascending or descending order.

   b) Find the median.

3. Refer to the frequency distribution of the following set of weights in kilogram.

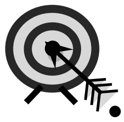
Weights (X)	Frequency
39	2
40	2
41	1
42	4
43	1
44	1
45	2
46	1
48	1
49	3
	N = 18

Find the median of the set of data.

CORRECT YOUR WORK. ANSWERS ARE AT THE END OF TOPIC 3.

Lesson 16: Median of Grouped Data

You learnt to calculate te median of ungrouped data using formula in the previous lesson.



In this lesson, you will:

- define median of grouped data
- use the formula to work out the median of grouped data

You learnt to work out the median of ungrouped data using formula in the previous lesson.

For grouped data, finding the median is more difficult. It cannot be found exactly but is estimated using interpolation.

Interpolation is the method of constructing new data points within the range of a discrete set of known data points.

For example:

Let us find the median of the scores in a revision test presented on Table 14.3 in Lesson 14.

SCORES IN A REVISION TEST

Class Interval	Frequency(f)	Class Limits	Cumulative Frequency	
		Lower - Upper	<cf	>cf
61-65	2	60.5 - 65.5	2	42
66-70	3	65.5 - 70.5	5	40
71-75	7	70..5 - 75.5	12	37
76-80	9	75.5 - 80.5	21	30
81-85	10	80..5 - 85.5	31	21
86-90	6	85.5 - 90.5	37	11
91-95	4	90.5 – 95.5	41	5
96-100	1	95.5 - 100.5	42	1
	N = 42			

When the data is given in a frequency distribution form as shown above, we first find out in what class interval we find the $\left(\frac{n}{2}\right)^{th}$ case. Proceeding from the small to larger values, we interpolate within the interval to determine the point that fulfils the condition in the definition of median. The table shows a frequency distribution of 42 scores. Half the scores (e.g. $\left(\frac{N}{2}\right)$ = 21) should lie above the median and half below. On the table, the class interval where $\left(\frac{N}{2}\right)$ = 21 falls is 76 - 80 whose exact lower limit is 75.5 and the upper limit is 80.5. This class interval is called the **median class.** Counting frequencies downward from the top to the interval 71 – 75 are 12 cases. To make 21 we need 9 out of the 9 cases in the class 76-80. Since we do not know exactly how the frequencies are distributed to an interval, we make the assumption that the number of cases within an interval are evenly distributed or spread over the distance from the lower limit to the upper limit of the class. In our example, 9 cases are evenly distributed from 75.5 to 80.5. To find how far above 75.5 we need to go in order to include the 12 cases we need below the median, we must go $\frac{9}{9}$ of way. Since the total distance or the length of the interval (class size) is 5 units, we therefore, go $\frac{9}{9}$ of 5 or exactly 5 units. Adding this to the lower limit of the class, we have 75.5 + 5 = 80.5 is the median.

To solve for the median of the class interval on page 118, the following steps are used.

1. Compute the less than cumulative frequencies.
2. Find $\frac{N}{2}$.
3. Locate the class interval in which the median class falls, and determine the exact lower limits of this interval.
4. Substitute the given values in the formula.

The formula we used to find the median of grouped data is:

$$\boxed{\mathbf{Mdn = L + \frac{\frac{N}{2} - F_b}{f_m}(i)}}$$

Median from below

Where: L = exact lower limit of the class interval where the median lies or median class

F_b = cumulative frequency below median class

f_m = frequency of the median class

N = size of the distribution

i = class size

The median can also be obtained by counting up from the bottom of the distribution until $\frac{N}{2}$ of the cases are included. In our example, the sum of the frequencies from the bottom up to and including the class interval 76 – 80 is 21. We do not need any of the next group of 9 cases to make 21. We, therefore, take the upper limit 80.5 as the median which checks with the values obtained when we count up from the bottom.

The formula we used is

$$\boxed{\mathbf{Mdn = U - \frac{\frac{N}{2} - F_a}{f_m}(i)}}$$

Median from above

Where: U = exact upper limit of the class interval where the median lies or median class

F_a = cumulative frequency above median class

f_m = frequency of the median class

N = size of the distribution or total number of cases

i = class size

Now, see the calculation using the two formulas.

From the example we have the following:

$\frac{N}{2} = 21$ ⟶ half the scores

L = 75.5 ⟶ exact lower limit of the median class

U = 80.5 ⟶ exact upper limit of the median class

i = 5 ⟶ class size

$F_b = 12$ ⟶ cumulative frequency below L

$F_a = 21$ ⟶ cumulative frequency above U

$f_m = 9$ ⟶ frequency of the median class

Solution: Substitute the given values in the formulas.

(1) $\text{Mdn} = L + \frac{\frac{N}{2} - F_b}{f_m}(i)$

$\text{Mdn} = 75.5 + \frac{21 - 12}{9}(5)$

$= 75.5 + \frac{9}{9}(5)$

$= 75.5 + 5$

$= 80.5$

(2) $\text{Mdn} = U - \frac{\frac{N}{2} - F_a}{f_m}(i)$

$\text{Mdn} = 80.5 - \frac{21 - 21}{9}(5)$

$= 80.5 - \frac{0}{9}(5)$

$= 80.5 - 0$

$= 80.5$

Therefore, the median is 80.5.

Example 2

Below is a computation of the median from the frequency distribution of scores in a Science Test.

SCORES IN A SCIENCE TEST

Class Interval	Frequency(f)	Class Limits		Cumulative Frequency	
		Lower	Upper	<cf	>cf
80-84	2	79.5 - 84.5		38	2
75-79	1	74.5 - 79.5		36	3
70-74	3	69.5 - 74.5		35	6
65-69	9	64.5 - 69.5		32	15
60-64	10	59.5 - 64.5		23	25
55-59	7	55.5 - 59.5		13	32
50-54	4	49.5 - 54.5		6	36
45-49	1	44.5 - 49.5		2	37
40-44	1	39.5 - 44.5		1	38
	N = 38				

To find the median:

1. Using the formula $Mdn = L + \dfrac{\frac{N}{2} - F_b}{f_m}(i)$

$$Mdn = 59.5 + \frac{19 - 13}{10}(5)$$

$$= 59.5 + \frac{6}{10}(5)$$

$$= 59.5 + 3$$

$$\mathbf{= 62.5}$$

2. Using the formula $Mdn = U - \dfrac{\frac{N}{2} - F_a}{f_m}(i)$

$$Mdn = 64.5 - \frac{19 - 15}{10}(5)$$

$$= 64.5 - \frac{4}{10}(5)$$

$$= 64.5 - 2$$

$$\mathbf{= 62.5}$$

NOW DO PRACTICE EXERCISE 16

Practice Exercise 16

1. Complete the Frequency Distribution Table below by filling in the empty column.

Class Interval	Frequency(f)	Class Limits		Cumulative Frequency	
		Lower	Upper	<cf	>cf
25 – 29	3				
30 – 34	2				
35 – 39	5				
40 – 44	8				
45 – 49	8				
50 – 54	8				
55 – 59	9				
60 – 64	6				
65 – 69	6				
70 – 74	3				
75 – 79	3				
80 - 84	3				
	N =				

2. Using the distribution in Question 1 answer the following:

A. Find the following values from the distribution above.

a) $\frac{N}{2}$

b) i

c) L

d) F_b

e) F_a

f) f_m

B. Compute the median

3. The following is a frequency distribution of examination marks.

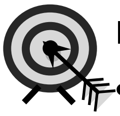
Class Interval	Frequency(f)
40 – 44	3
45 – 49	3
50 – 54	4
55 – 59	6
60 – 64	6
65 – 69	14
70 – 74	9
75 – 79	8
80 – 84	4
85 – 89	2
90- 94	1
	N = 60

Compute the median.

CORRECT YOUR WORK. ANSWERS ARE AT THE END OF TOPIC 3

Lesson 17: Mode

You learnt how to find the mean and the median of ungrouped and grouped data in the previous lessons.



In this lesson, you will:

- define mode of ungrouped and grouped data
- use the formula to work out the mode of ungrouped and grouped data

As we have already learnt, when a set of data is given, there are three numbers which give us the important information about it. These are the mean, median and mode.

Mode for Ungrouped Data

For ungrouped data, the mode is defined as that datum value or specific score which has the highest frequency. It is the most frequently occurring score or loosely speaking, the most popular score.

Example 1

Find the mode of the following data:

9, 9, 11, 11, 11, 13, 13, 13, 14, 15, 15, 15, 17, 17, 17, 17, 18

By inspection, the mode is 17 because it appears the most number of times.

Example 2

The ages of audience members at a rap concert were recorded. The results were listed below.

12, 12, 14, 14, 12, 15, 16, 11, 15, 13, 14, 15

16, 16, 14, 16, 14, 16, 13, 13, 13, 13, 14, 15

Find the mode.

Solution: Arrange the data in order from lowest to highest in a frequency table.

Ages	Frequency
11	1
12	3
13	5
14	6
15	4
16	5
	N= 24

The mode is 14.

Example 3

To pass a Typing class, the students need to have a typing speed of 30 words a minute. In a test the results were:

36	45	43	32	29	28
37	34	38	29	31	32
34	38	39	34	36	32
35	41	36	35	32	31
34	35	34	36	37	39

a) What was the best typing speed?
b) What was the worst typing speed?
c) What was the modal score?

Answers: a) The best typing speed was 45 words per minute. It is the highest typing speed.

b) The worst typing speed was 28 words per minute. It is the lowest typing speed.

c) The modal score was 34 words per minute. It is the typing speed that appears the most.

Mode for Grouped Data

You have seen that by just observing the given ungrouped data carefully its mode can be obtained. However, for grouped data it is not possible to find the mode just by observation.

The first step towards finding the mode for grouped data is to locate the class interval with the maximum or highest frequency. The class interval corresponding to the maximum or highest frequency is called the **modal class.** The mode of this data lies in between this data and is calculated using the formula

$$\boxed{\text{Mode} = L + \frac{F_1 - F_0}{2F_1 - F_0 - F_2}(i)}$$

Where: L = exact lower limit of the class interval where the mode lies or modal class

F_1 = frequency of the modal class

F_2 = frequency after modal class

F_0 = frequency before the modal class

i = class size

Let us understand this method more clearly with the help of an example.

See example on the next page.

Example

Find the mode for data below.

Class Interval	Frequency
25-29	3
30-34	2
35-39	5
40-44	8
45-49	8
50-54	8
55-59	9
60-64	6
65-69	6
70-74	3
75-79	3
80-84	3
	N= 64

Solution: First let us locate the modal class.

As you can see in the distribution, the modal class is the class interval 55 – 59 with 9 as the highest frequency (shaded part of the distribution table). This means that the mode lies in this class interval.

We can now outline the following data:

L = 55, exact lower limit of the modal class
F_1 = 9, frequency of the modal class
F_2 = 6, frequency after the modal class
F_0 = 8, frequency before the modal class
i = 5, class size

Now to calculate the Mode, substitute all of these values in the formula:

$$\text{Mode} = L + \frac{F_1 - F_0}{2F_1 - F_0 - F_2}(i)$$

Thus we have,

$$\text{Mode} = 55 + \left[\frac{9-8}{2(9)-8-6}\right](5)$$

$$= 55 + \left[\frac{1}{18-14}\right](5)$$

$$= 55 + \left[\frac{1}{4}\right](5)$$

$$= 55 + 1.25$$

$$=56.25$$

The mode of the distribution is 56.25.

Example 2

The following is a frequency distribution of an entrance examination. Find the mode of the scores.

Class Interval	Frequency
40-44	7
45-49	10
50-54	14
55-59	17
60-64	19
65-69	26
70-74	20
75-79	18
80-84	13
85-89	9
90-94	7
	N= 160

Solution:

As you can see in the distribution, the modal class is the interval 65-69 because it has the largest frequency which is 26 and the mode lies in of this class interval.

We can now outline the following data:

L = 65, exact lower limit of the modal class
F_1 = 26, frequency of the modal class
F_2 = 20, frequency after the modal class
F_0 = 19, frequency before the modal class
i = 5, class size

Now to calculate the Mode, substitute all of these values in the formula:

$$\text{Mode} = L + \frac{F_1 - F_0}{2F_1 - F_0 - F_2}(i)$$

Thus we have,

$$\text{Mode} = 65 + \left[\frac{26 - 19}{2(26) - 19 - 20}\right](5)$$

$$= 65 + \left[\frac{7}{52 - 39}\right](5)$$

$$= 65 + \frac{35}{13}$$

$$= 65 + 2.7$$

$$= 67.7$$

The mode of the distribution is 67.7.

NOW DO PRACTICE EXERCISE 17

Practice Exercise 17

1. Find the mode of each of the following set of data.

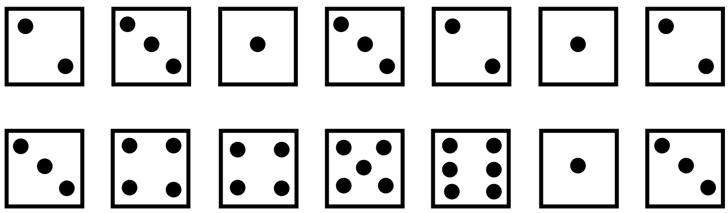
   a) 1, 3, 4, 2, 3, 4, 2, 5, 3, 2, 5
   b) 1, 3, 2,3, 1, 3, 2, 2, 1, 3, 2, 3
   c) 5, 7, 5, 6, 7, 6, 5, 6, 5, 6, 5,6, 7, 7, 7, 7
   d) 5, 3, 4, 5, 6, 3, 5, 4, 3, 6, 4, 5, 6

2. In various shops, a packet of beans was priced in kina as follows:

   K18, K19, K19, K21, K18, K21, K 18, K23, K18, K23, K23

   What is the modal price?

3. A die was thrown 14 times as follows:



   What was the modal score?

Refer to the frequency distribution table below to answer Question 4 and 5.

Class Interval	Frequency
5 - 25	12
26 - 45	8
46 - 65	14
66 - 85	20
86 -105	6

4. Find the following values in the distribution.

   a) i
   b) L
   c) F_1
   d) F_2
   e) F_0

5. Compute the mode.

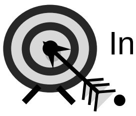
6. Find the mode for the following grouped data.

Class Interval	Frequency
1 - 5	3
6 - 10	7
11 - 15	12
16 - 20	5
21 -25	3

CORRECT YOUR WORK. ANSWERS ARE AT THE END OF TOPIC 3

Lesson 18: Mixed Problems

You learnt to work out the mean, mode and median of grouped and ungrouped data using formula in the previous lessons.



In this lesson, you will:

- solve mixed problems involving measures of central tendency.

People in many situations use the measures of central tendency or location in order to solve problems and make informed decisions. For examples, in selecting the type of product consumers will buy, collecting and grading students, making decisions about the most appropriate crop from a particular type of plant and so on.

You will need your skills on the different measures of central tendency or averages (mean, median and mode) to solve problems in this lesson.

Here are some examples.

Example 1

The weekly earnings of 10 employees of an insurance company taken at random are as follows; K450, K500, K525, K550, K575, K580, K600,K630, K640 and k700.

What is the weekly mean earning of the 10 employees?

Solution; Recall the formula for the mean of ungrouped data.

$$\overline{X} = \frac{\sum x}{N}$$

Where: $\sum x$ = the sum of all the given values.

N = number of values

Step 1 Find $\sum x$ by adding all the values.

$\sum x$ = K450 + K500 + … + K700

= 5750

Hence, $\overline{X} = \frac{K5750}{10}$

= K575

Therefore, the weekly mean earning of the 10 employees is K575.

Example 2

In order to keep track of the products in the warehouse, the storekeeper records the number of items sold per day every week.

(a) Here is the record of the number of boxes of bars of chocolate sold during the first week of August.

Monday	24
Tuesday	31
Wednesday	27
Thursday	26
Friday	28
Saturday	29
Sunday	33

What is the median?

Solution:

Arrange the seven values in order from least to greatest.

Hence, we have 24 26 27 28 29 31 33

Since the number of values is odd, the median he median is the value in the middle position. In the distribution, 28 is the value in the middle position.

Therefore, the median is 28.

Example 3

Hennie and Helen own a clothes shop and they keep a record of their sales. They want to know their average daily takings.

The tables below show their takings over a fortnight.

FIRST WEEK:

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
K130	K130	K129	K106	K96	K594

SECOND WEEK:

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
K125	K130	K110	K132	K118	K468

There are three types of averages: the mean, the median and the mode.

First, find the mode.

As you know, the mode is the value or the score that appears most often.

In the tables of the shop's takings K130 appears most often.

Therefore, the mode is K130.

Now let us find the Median.

As you know, the median is the middle value when the data is arrange in numerical order.

When written in numerical order, the numbers in the tables are:

K96, K106, K110, K118, K125, K129, K130, K130, K130, K132, K468, K594

Notice that there are 12 items, so n = 12. Use the formula

$$\text{Median} = X_{\left(\frac{n+1}{2}\right)}$$

$$\text{Median} = X_{\left(\frac{12+1}{2}\right)}$$

$$= X_{\left(\frac{13}{2}\right)}$$

$$= X_{6.5}$$

This means that the median is value $\frac{1}{2}$ way between the 6^{th} and 7^{th} values.

Therefore the median is $\frac{129+130}{2}$ = **K129.5.**

Now let us work out the third average which is the mean.

To find the mean of the values on the clothes shop, we add all the data and divide the total by the number of items or use the formula: $\overline{X} = \frac{\Sigma fX}{N}$

$$\text{Hence,} \quad \overline{X} = \frac{96+106+110+...+594}{12}$$

$$\overline{X} = \frac{2268}{12}$$

$$\overline{X} = 189$$

Therefore the mean is **K189.**

Example 4

Find the (a) mean
(b) mode

for the following grouped data shown in the table below:

Class Intervals	**Frequency(f)**
0 – 3	*2*
4 – 7	*3*
8 – 11	*4*
12 – 15	*9*
16 - 19	*2*

Solution:

(a) Finding the Mean.

Class Intervals		**Class Centres or Midpoints(M)**	**Frequency(f)**	**fM**
LS	**HS**			
0 – 3		1.5	*2*	3
4 – 7		5.5	*3*	16.5
8 – 11		9.5	*4*	38
12 – 15		13.5	*9*	121.5
16 - 19		17.5	*2*	35
			N = 20	ΣfM = 214

Step 1 Determine the midpoints of each class interval. Use the formula:

$$M = \frac{LS + HS}{2}$$

e.g. $\frac{3+0}{2} = \frac{3}{2} = \mathbf{1.5}$**;** $\frac{7+4}{2} = \frac{11}{2} = \mathbf{5.5}$; and so on.

Step 2 Get the product of each midpoint and the corresponding frequency within its interval to obtain ΣfM.

Step 3 Apply the formula by substituting the values ΣfM and N.

$$\text{Mean} = \frac{\sum fM}{N} = \frac{214}{20} = 10.7$$

Therefore the mean is 10.7.

(b) The modal class is the class interval 12 – 15 having the largest frequency.
Class size (i) = 4

Find the Mode using the formula: Mode = L + $\frac{F_1 - F_0}{2F_1 - F_0 - F_2}(i)$

Thus we have,

$$\text{Mode} = 12 + \left[\frac{9-4}{2(9)-4-2}\right](4)$$

$$= 12 + \left[\frac{5}{18-6}\right](4)$$

$$= 12 + \frac{20}{12}$$

$$= 12 + 1.7$$

$$= 13.7$$

The mode of the distribution is 13.7.

NOW DO PRACTICE EXERCISE 18

Practice Exercise 18

Solve the following problems.

1. The operating expenses of a canteen for four weeks are as follows:

First Week:	K1500
Second Week:	K1450
Third Week:	K1400
Fourth week:	K1500

What is the weekly mean operating expense of the canteen?

2. The duration in minutes of telephone calls in a pay telephone on a certain day was:

3, 4, 5, 6, 7, 8, 10, 12, 15, 16, 17

Find the median.

3. The mean height of of a group of eight students is 165 cm.

(a) What is the total height of all the students?

If one student whose height is 168 cm joins the group,

(b) What would be the mean height of all the students?

4. The weekly pocket money of a group of students is recorded below:

K10, K8, K8, K5.50, K3, K5. K4, K7.50, K5

(a) What is the mode?

(b) Put the amounts in order and find the median.

(c) Calculate the mean weekly pocket money.

---

5. Below is a table showing the grouped distribution of 30 scores in a Maths test.

Scores	**Frequency (f)**	**Midpoint (M)**	**Product (fM)**
95-99	1	97	97
90-94	3	92	276
85-89	4	87	348
80.84	6	----	-----
75-79	5	----	-----
70-74	4	----	-----
65-69	2	----	-----
60-64	3	----	-----
55-59	1	----	-----
50-54	1	---	-----
	N = 30		∑fM = ____

(a) Complete the blank spaces on the table.

(b) Find the mean.

CORRECT YOUR WORK. ANSWERS ARE AT THE END OF TOPIC 3

TOPIC 3: SUMMARY

This summarizes some of the important ideas and concepts to remember.

- The **Mean** is the arithmetic average of a set of data.
- The **mean of ungrouped data** is calculated by using the formula:

$$\boxed{\overline{X} = \frac{\sum fX}{N}}$$

- The **mean of grouped data** is calculated by using the formula:

$$\boxed{\overline{X} = \frac{\sum fM}{N}}$$

- The **Median** is the middle value in an ordered set of data.
- The **median of ungrouped data** is the $\left[\frac{n+1}{2}\right]^{th}$ value. The median is the value in the middle position if the number of values is **odd**.If the number of values is **even**, the median is the average of the two values in the middle position.
- The **median for grouped data** is calculated using the formula:

$$\boxed{\mathbf{Mdn} = \mathbf{L} + \frac{\frac{N}{2} - F_b}{f_m}(i)} \quad \text{or} \quad \boxed{\mathbf{Mdn} = \mathbf{U} - \frac{\frac{N}{2} - F_a}{f_m}(i)}$$

- The **Mode** of a set of numbers or data is that value which occurs with the greatest frequency or most often.

- The **mode for grouped data** is the midpoint of the class interval with the largest frequency. it is calculated by using the formula

$$\boxed{\mathbf{Mode} = \mathbf{L} + \frac{F_1 - F_0}{2F_1 - F_0 - F_2}(i)}$$

REVISE LESSONS 13-18. THEN DO TOPIC TEST 3 IN ASSIGNMENT BOOK 3.

ANSWERS TO PRACTICE EXERCISES 13-18

Practice Exercise 13

1. (a) 87 (b) 24.67 (c) 70.4

2. 175 km

3. 80.91

4. 5.2 km

5. (a) 18 15 14 13 12 11 10 9 8

(b)

Days (X)	Frequency (f)	(f)(X)
18	**1**	**18**
15	**1**	**15**
14	**1**	**14**
13	**1**	**13**
12	**2**	**24**
11	**1**	**11**
10	**1**	**10**
9	**2**	**18**
8	**2**	**16**
	N = 12	**ΣfX = 139**

(c) 11.58

Practice Exercise 14

1. (a)

Class Interval	Frequency(f)	Midpoints(M)	fM
39-40	4	39.5	158
41-42	5	41.5	207.5
43-44	2	43.5	87
45-46	3	45.5	136.5
47-48	1	47.5	47.5
49-50	3	49.5	148.5
	N = 18		ΣfM = 785

(c) 43.61

2. (a)

Class Interval	Frequency(f)	Midpoints(M)	fM
3 – 5	3	4	12
6 – 8	4	7	28
9 – 11	6	10	60
12 – 14	6	13	78
15 – 17	6	16	96
18 - 20	5	19	95
	N = 30		ΣfM = 369

(b) 12.3

Practice Exercise 15

1. (a) 35 (b) 75 (c) 105.5

2. (a) 121, 169, 198, 201, 219, 232, 251, 256, 283, 291, 305, 343, 346, 462

(b) 253.5

3. 42.5

Practice Exercise 16

1.

Class Interval	Frequency(f)	Class Limits		Cumulative Frequency	
		Lower	Upper	<cf	>cf
25 – 29	3	**24.5**	**29.5**	**3**	**64**
30 – 34	2	**29.5**	**34.5**	**5**	**61**
35 – 39	5	**34.5**	**39.5**	**10**	**59**
40 – 44	8	**39.5**	**44.5**	**18**	**54**
45 – 49	8	**44.5**	**49.5**	**26**	**46**
50 – 54	8	**49.5**	**54.5**	**34**	**38**
55 – 59	9	**54.5**	**59.5**	**43**	**30**
60 – 64	6	**59.5**	**64.5**	**49**	**21**
65 – 69	6	**64.5**	**69.5**	**55**	**15**
70 – 74	3	**69.5**	**74.5**	**58**	**9**
75 – 79	3	**74.5**	**795**	**61**	**6**
80 - 84	3	**79.5**	**84.5**	**64**	**3**
	N = **64**				

2. A. (a) 32
(b) 5
(c) 49.5
(d) 26
(e) 30
(g) 8

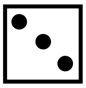
B. 53.5

3. 67.35

Practice Exercise 17

1. (a) 2 and 3
(b) 3
(c) 7
(d) 5

2. K18

3. 

4. (a) 20 (b) 66 (c) 20 (d) 6 (e) 14

5. 72

6. 13.08

Practice Exercise 18

1. K1462.50

2. 8 min

3. (a) 1320 cm

(b) 165.33 cm

4. (a) K5 and K8, the distribution is bimodal

(b) K3, K4, K5, K5, K5.50, K7.50, K8, K8, K10; $\overline{X}$ = K5.5

(c) K6.20

5. (a)

Scores	**Frequency(f)**	**Midpoint(M)**	**Product (fM)**
95-99	1	97	97
90-94	3	92	276
85-89	4	87	348
80-84	6	**82**	**492**
75-79	5	**77**	**385**
70-74	4	**72**	**288**
65-69	2	**67**	**134**
60-64	3	**62**	**186**
55-59	1	**57**	**57**
50-54	1	**52**	**52**
	N = 30		∑fM = **2315**

(b) $\overline{X} = 77.17$

END OF TOPIC 3

TOPIC 4

MEASURES OF SPREAD

TOPIC 4: MEASURES OF SPREAD

Statistical averages give us some idea about the magnitude of the data or quantities in the distribution, but it tells us nothing about the spread of the distribution.

This topic will give you an idea of how the data in the distribution are dispersed or are spread.

There are four measures of spread or dispersion which are used and these are the range, the interquartile range or (IQR), the Semi-inter-quartile range or Variance and the Standard deviation.

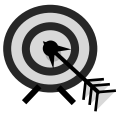
In this topic, only the first one will be discussed. The other two will be discussed in your higher mathematics.

In this topic, you will:

- define the range of ungrouped and grouped data
- calculate the range of ungrouped data using the formula highest score minus lowest score.
- calculate the range of grouped data using the formula highest class limit minus lowest class limit.

Lesson 19: Range of Ungrouped Data

You've learnt the different measures of tendency or location in the last topic.



In this lesson, you will:

- .define the range of ungrouped data.
- calculate the range of ungrouped data.

A set of numbers may be summed up and a single number is computed to represent the whole set. This is the measure of central tendency or location. However, as a descriptive measure, this measure is incomplete.

Knowing only the measures of central tendency does not give us a complete picture of the characteristics of the data distribution. In other words, it is not enough to simply have the average or the median of a set of data. We also need a value that will disclose how closely or how widely scattered these variables are from the mean. There is the need also to compute a measure of the range, the scattering, fluctuation, spread, dispersion or variability of the scores within the set.

The simplest measure of dispersion is the range.

The range is the difference between the largest and the smallest observations.

For ungrouped data, if you subtract the lowest score from the highest score, you get the range.

The formula is:

Range = Highest score – Lowest Score

Example 1

Find the range of the distribution if the highest score is 120 and the lowest score is 21.

Solution:

Range = Highest Score – Lowest Score

= 120 – 21

= 99

Example 2

The temperature in ºC was recorded at 2-hourly intervals at a location in the desert. These are the results:

-4, -12, -2, 5, 20, 27, 25, 32, 38, 39, 27.

Find the range.

Solution: Arrange the numbers in order.

-12, -4, -2, 5, 20, 25, 27, 27, 32, 38, 39

Range = Highest score – Lowest score

= 39 – (-12)

= 51

Therefore the range is 51ºC.

Example 3

Find the range for the following data.

Scores	Frequency
50	3
51	5
52	8
53	6
54	2
55	4

Solution:

Highest Score = 55
Lowest Score = 50

Range = Highest Score – Lowest score

= 55 – 50

= 5

The range, like the mode, is a very unstable measure in statistics. It can vary from sample to sample. The range can be used justifiably, however, when you want a quick measure of variability and you do not have time to compute other measures of variability.

Here are other examples of finding the range.

Example 4

The marks obtained in a test, by two sets of students are given in the following table.

Boys	40	50	46	52	46	51	85
Girls	37	72	39	68	48	74	73

Find the range of: (a) the boys' marks
(b) the girls' marks.

Solution: (a) Range of Boys' marks

Range = Highest Score – Lowest Score

= 85 – 40

= 45

(b) Range of Girls' marks

Range = Highest score – Lowest score

= 74 – 37

= 37

Example 5

The results of a year 9 test are:

100 77 93 87 93 40 73 27 100 89 100 87
87 100 100 83 93 100 83 74 89 81 52 94

What is the range for the test results?

Solution: The highest score is 100.

The lowest score is 27.

Range = Highest score – Lowest score

= 100 – 27

= 73

NOW DO PRACTICE EXERCISE 19

PRACTICE EXERCISE 19

1. Find the range of the scores given:

   (a) 20, 20, 20, 23, 25, 27

   (b) 11, 13, 13, 16, 170

   (c) 2, 3, 3, 4, 5, 6, 7, 8, 9

   (d) 27, 28, 29, 27, 30, 31, 27, 31, 30

   (e) 51, 52, 54, 55, 57, 57, 58, 59

---

2. The weight in kilograms of students in a certain class is listed below.

   48, 42, 42, 45, 42, 40
   39, 43, 49, 49, 49, 42
   39, 45, 44, 41, 40, 46

   Find the range of the distribution.

---

3. The marks in Mathematics and Science for ten students are shown below.

Mathematics	50	60	65	70	72	74	78	78	80	81
Science	42	53	58	64	66	64	71	70	71	75

Find the range of: (a) the marks in Mathematics
(b) the marks in Science.

4. The marks scored by Jackson and Mac in eight topic test in Mathematics are shown below.

Test	1	2	3	4	5	6	7	8
Jackson	82	81	91	84	82	75	88	54
Mac	81	80	86	83	88	72	86	79

(a) Find the total marks scored by each student.

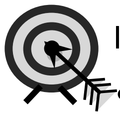
(b) In how many tests did Mac score more marks than Jackson?

(c) Find the range of each student's scores.

CORRECT YOUR WORK. ANSWERS ARE AT THE END OF TOPIC 4.

Lesson 20: Range of a Grouped Data

You have learnt to find the range of an ungrouped data.



In this lesson, you will:

- define the range of grouped data.
- calculate the range of grouped data.

When data is presented in a frequency distribution, wherein the items in a set of data are arranged into groups or classes and the number of classes occurring in each group is indicated, they are called **grouped data.**

To find the range for a frequency distribution, just get the difference between the upper limit of the highest class interval and the lower limit of the lowest class interval.

The grouped data formula for range is:

Range = Highest Class Upper Limit – Lowest Class Lower Limit

Example 1

Find the range for the frequency distribution shown below.

Class Interval	Frequency
90-94	4
95-99	6
100-104	10
105-109	13
110-114	8
115-119	6
120-124	3
	N=50

Solution:

The Highest Class Interval is 120 – 124 so the highest class upper limit is 124.5.
The Lowest Class Interval is 90 – 94, so the lowest class lower limit is 89.5.

Using the formula: Range = Highest Class Upper Limit – Lowest Class Lower Limit

= 124.5 – 89.5

= 35

Therefore, the range of the grouped data is 35.

Example 2

Find the range for the following grouped data.

Class Interval	Frequency
46-50	1
41-45	4
36-40	8
31-35	12
26-30	10
21-25	11
	N=46

Solution:

The Highest Class Interval is 46 – 50 so the highest class upper limit is 50.5.
The Lowest Class Interval is 21 – 25, so the lowest class lower limit is 20.5.

Range = Highest Class Upper Limit – Lowest Class Lower Limit
= 50.5 – 20.5
= 30

Therefore, the range is 30.

Example 3

Below is the grouped frequency distribution of the scores of 42 students in a Mastery Test.

Class Interval	Frequency
96-100	1
91-95	4
86-90	6
81-85	10
76-80	9
71-75	7
66-70	3
61-65	2
	N=42

Find the range of scores.

Solution:

The Highest Class Interval is 96 – 100 so the highest class upper limit is 100.5.
The Lowest Class Interval is 61 – 65, so the lowest class lower limit is 60.5.

Range = Highest Class Upper Limit – Lowest Class Lower Limit
= 100.5 – 60.5
= 40

Therefore, the range is 40.

If the data has open-ended intervals, we use the same approach we have used throughout. Treat the open-ended interval as if it has the same width (size) as its adjacent interval.

Example 4

Below is the grouped frequency distribution of the number of videos purchased per week.

NUMBER OF VIDEOS PURCHASED PER WEEK

Class Interval	Frequency
Under 20	1
20-29	17
30-39	31
40-49	12
50 or over	2
	N=63

For this data, the lowest class interval is 'Under 20'. Treating this as having the same with as the class interval next to it, the class interval is assumed to be '10-19', so the lowest class lower limit is 9.5. Similarly, the highest class interval is '50 or over'. This is treated as '50-59' since the class interval adjacent to it has this width. The higher limit is 59.5.

Using the formula: Range = Highest Class Upper Limit – Lowest Class Lower Limit

Range = 59.5 – 9.5

= 50

Therefore, the range is 50.

The range has the following properties:

1. It is easy to understand.
2. It is easy to calculate.
3. It depends on the extreme values so is susceptible (subject) to odd result.
4. It only uses two values, the remaining data is ignored.
5. It is only rarely used for further statistical work.

NOW DO PRACTICE EXERCISE 20

Practice Exercise 20

1. A firm has recorded the number of applicants for posts it advertises. The figure are given in the following frequency table:

Applicants	Frequency
1 – 5	17
6 – 10	38
11 – 15	19
16 – 20	13
21 – 25	5
26 - 30	1
	N = 93

Find the range.

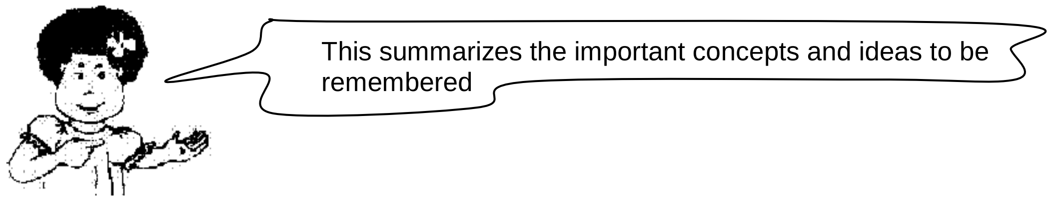
2. The lengths of steel bars gave the following frequency distribution:

Lengths (in m)	Frequency
Under 1.95	4
1.95 but less than 2.00	12
2.00 but less than 2.05	27
2.05 but less than 2.10	28
2.10 but less than 2.20	61
2.20 but less than 2.50	25
2.50 or over	3
	N = 160

Find the range.

CORRECT YOUR WORK. ANSWERS ARE AT THE END OF TOPIC 4.

TOPIC 4 SUMMARY



- **Measures of Spread** refer to the measures that disclose how closely or how widely scattered are the scores or variables in the distribution to the middle of the distribution or from the mean. Measures of spread are also known as the Measures of Variability or Measures of Dispersion.
- The terms **variability, spread and dispersion** are synonyms and refer to how spread the distribution of data is.
- The **Range** is the simplest measure of variability or spread to calculate. It is simply the highest score minus the lowest score.
- To find the range of ungrouped data, you subtract the lowest score from the highest score. Use the formula:

Range = Highest Score – Lowest Score

- To find the range of grouped data, you subtract the lowest class lower limit from the highest class higher limit. Use the formula:

Range = Highest class upper limit – Lowest class lower limit

REVISE LESSONS 13-18. THEN DO TOPIC TEST 4 IN ASSIGNMENT BOOK 3.

ANSWERS TO PRACTICE EXERCISES 19-20

Practice Exercise 19

1. (a) 7 (b) 159 (c) 7 (d) 4 (e) 8

2. Range = Highest Score – Lowest Score
   = 49 – 39
   = 10

3. (a) Marks in Mathematics: Range = H.S. – L.S.
   = 81 – 50
   = 31

   (b) Marks in Science: Range = H.S. – L.S.
   = 75 – 42
   = 33

4. (a) Total marks of Jackson = 637
   Total marks of Mac = 655

   (b) Mac scores more marks than Jackson twice (in the 5th and 8th test)

   (c) Range of Jackson's scores = H. S. – L.S.
   = 91 – 54
   = 37

   Range of Mac's scores = H.S. – L.S.
   = 88 – 72
   = 16

Practice Exercise 20

1. Range = Highest class higher limit – Lowest class lower limit
   = 30.5 – 0.5
   = 30

2. Range = Highest class higher limit – Lowest class lower limit
   = 2.80 – 1.90
   = 0.90 metres

END OF UNIT 3

REFERENCES

- Oxford, The resource for the new AQA specification: **Statistics GCSE for AQA** by: Jayne Kranat, Brian Housden and James Nicholson
- Oxford Mathematics; **Higher GCSE for AQA**, Linear Specification, Editors:Peter McGuire and Ken Smith
- Oxford Mathematics **Intermediate GCSE** by Sue Briggs, peter McGuire,Susan Shilton and Ken Smith
- infinity.cos.edu/faculty/woodbury/Stats/Tutorial/Data_Pop_Samp.htm
- http:/ninemsn.com.au/ stattrek.com/**statistics/datacollections-methods.**.aspx
- FODE Grade 9 Formal Mathematics Course Books 6
- NDOE Secondary School Mathematics 9A
- NDOE Secondary School Mathematics 10B
- Grade 9 Mathematics Outcome Based Edition
- Statistics: a Simplified Approach