NOUN

NATIONAL OPEN UNIVERSITY OF NIGERIA

INTRODUCTION TO QUANTITATIVE METHOD II

ECO 154

SCHOOL OF ARTS AND SOCIAL SCIENCES

COURSE GUIDE

Course Developer:

MAKU, Olukayode Emmanuel
School of Arts and Social Sciences,
National Open University, Nigeria.
14/16 Ahmadu Bello Way,
Victoria Island, Lagos.

Course Editor:
Dr. IBRAHIM, Adeolu
Bakare

CONTENTS

Introduction

Welcome to ECO 154: Introduction to Quantitative Method II
ECO 154: Introduction To Quantitative Method II is a three unit, one semester undergraduate course. It comprises 24 study units, subdivided into six modules. The materials have been developed with the Nigerian context in view by using simple and local examples. This course guide gives you an overview of the course. It also provides you with organization and the requirement of the course.

Course Aims

(a) To expose student to the basic definitions and scope of statistics.
(b) To identify and explain the basic concepts and branches of statistics.
(c) To improve students' knowledge about the various measures of central tendency.
(d) To acquaint students with the basic measures of dispersion and partition.
(e) To give students insight into the basic concepts of moment, skewness and kurtosis.
(f) To make students understand random variables and probability distribution.
(g) To introduce students to the meaning and components and importance of index number to economics, business and finance.

Course Objective

To achieve the aims above, we have some overall objectives. Each unit has its objectives. This will guide in your study. They are usually stated at the beginning of each unit; and when you are through with studying the units, go back and read the objectives. This would help you accomplish the task you have set to achieve. On completion of the course, you should be able to: -

(a) Define and explain the concept of statistics;
(b) Identify the various ways of collecting and organizing data;
(c) Demonstrate adequate skills in computing the measures of central tendency for grouped and ungrouped data;
(d) Compute and interpret the measures of dispersion for both grouped and ungrouped data;
(e) Compute and interpret moments, skewness and kurtosis of a set of observations;
(f) Calculate simple experimental probability and the probabilities in random distribution; and
(g) Calculate and interpret various components of index numbers.

Working through the Course

To complete this course you are required to go through the study units and other related materials. You will also need to undertake practical exercises for which you need a pen, a notebook and other materials that will be listed in this guide. The exercises are to aid you in understanding the basic concept and principles being taught in this course.

At the end of each unit, you will be required to submit written assignments for assessment purpose. At the end of the course, you will write a final examination.

Course Materials

The major materials you will need for this course are:

1. Course guide.
2. Study units.
3. Relevant textbooks, the one listed under each unit.
4. Assignment file.
5. Presentation schedule.

Study Units

There are 24 units in this course as follows: -

Module 1: Basic Introduction
Unit 1: Meaning and Relevance of Statistics
Unit 2: Types/Branches of Statistics
Unit 3: Basic Concepts in Statistics
Unit 4: Collection of Data
Unit 5: Organization of Data

Module 2: Representation o Data
Unit 1: Tables
Unit 2: Graphs
Unit 3: Charts
Unit 4: Histogram and Curves

Module 3: Basic Statistical Measures of Estimates
Unit 1: Measures of Central Tendency Ungrouped Data
Unit 2: Measures of Central Tendency of Grouped Data
Unit 3: Measures of Dispersion
Unit 4: Measures of Partition

Module 4: Moment, Skewness and Kurtosis
Unit 1: Moments
Unit 2: Skewness
Unit 3: Kurtosis

Module 5: Basic Statistical Measures of Estimates
Unit 1: Basic Concept in Probability
Unit 2: Use of Diagram in Probability
Unit 3: Experimental Probability Rules
Unit 4: Experimental Probability
Unit 5: Random Variable and Mathematics of Expectation

Module 6: Index Numbers
Unit 1: Meaning and Scope of Price Index Number
Unit 2: Weighted Index Number
Unit 3: Other Index Number

Textbooks and References

Certain books are recommended in the course. You may wish to purchase or access them for further reading or practices

Assignment File

An assignment file and a marking scheme will be made available to you. In this file, you will find all the details of the work, you must submit to your tutor for marking. The marks you obtain from these assignments will count towards the final mark you obtain for this course. Further information on assignment will be found in the assignment file itself and later in the Study Guide in the section on assessment.

Tutor-Marked Assignments (TMAs)

You will need to submit a specific number of Tutor-Marked Assignment (TMAs). Every unit in this course has a tutor-marked assignment. You will be assessed on four of them, but the best three marks for the best three (that is the highest three of the four marked) will be counted. The total marks for the best three assignments will be 30 percent of your total work.

Assignment questions for the unit in this course are contained in the Assignment File. When you have completed each assignment, send it, together with the Tutor-Marked Assignment (TMA) form to your tutor. Make sure each assignment reach your tutor on or before the deadline for submission. If for any reason, you cannot complete your work on time, contact your tutor to discuss the possibility of an extension. Extension will not be granted after due date, unless under exceptional circumstances.

Final Examination and Grading

The final examination of ECO 154 will be of three hours' duration. All areas of the course will be examined. Find time to study and revise the unit all over before your examination. The final examination will attract 70 percent of the total course grade. The examination shall consist of questions which reflect the type of self-testing practice exercise and tutor-marked assignment you have previously come across. All areas of course will be assessed. You are advised to revise the entire course after studying the last unit before you sit for the examination. You will also find it useful to review your tutor-marked assignments and the comments of your tutor on them before the final examination.

Course Marking Scheme

The following table lays out the actual course marks allocation is broken down.

Assessment	Marks
Assignment (Best three assignment out of the four marked)	30%
Final Examination	70%
Total	100%

Presentation Schedule

The dates for submission of all assignments will be communicated to you. You will also be told the date for the completion of the study units and dates of the examination.

Course Overview

The table below brings together the units, number of weeks you should take to complete them and the assignments that follow them.

Unit	Title of Work	Week's Activity	Assessment (End of Unit)
	Course Guide		
	Module 1: Basic Introduction		
1	Unit 1: Meaning and Relevance of Statistics	Week 1	Assignment 1
2	Unit 2: Types/Branches of Statistics	Week 1	Assignment 2
3	Unit 3: Basic Concepts in Statistics	Week 2	Assignment 3
4	Unit 4: Collection of Data	Week 2	Assignment 4
5	Unit 5: Organization of Data	Week 3	Assignment 5
	Module 2: Representation o Data		
1	Unit 1: Tables	Week 4	Assignment 1
2	Unit 2: Graphs	Week 4	Assignment 2
3	Unit 3: Charts	Week 5	Assignment 3
4	Unit 4: Histogram and Curves	Week 5	Assignment 4
	Module 3: Basic Statistical Measures of Estimates		
1	Unit 1: Measures of Central Tendency Ungrouped Data	Week 6	Assignment 1
2	Unit 2: Measures of Central Tendency of Grouped Data	Week 6	Assignment 2
3	Unit 3: Measures of Dispersion	Week 7	Assignment 3
4	Unit 4: Measures of Partition	Week 7	Assignment 4
	Module 4: Moment, Skewness and Kurtosis		
1	Unit 1: Moments	Week 8	Assignment 1
2	Unit 2: Skewness	Week 9	Assignment 2
3	Unit 3: Kurtosis	Week 10	Assignment 3
	Module 5: Basic Statistical Measures of Estimates		
1	Unit 1: Basic Concept in Probability	Week 11	Assignment 1
2	Unit 2: Use of Diagram in Probability	Week 12	Assignment 2
3	Unit 3: Experimental Probability Rules	Week 12	Assignment 3
4	Unit 4: Experimental Probability	Week 13	Assignment 4
5	Unit 5: Random Variable and Mathematics of Expectation	Week 13	Assignment 5
	Module 6: Index Numbers		
1	Unit 1: Meaning and Scope of Price Index Number	Week 14	Assignment 1
2	Unit 2: Weighted Index Number	Week 14	Assignment 2
3	Unit 3: Other Index Number	Week 15	Assignment 3
	Revision	Week 16	
	Examination	Week 17	
	Total	17 Weeks	

How to Get the Most from this Course

In distance learning, the study units replace the university classroom lectures. This is one of the merits of distance learning; you can read and work through the outlined study materials at your own pace, time and place of your choice. It's all about the conception that you are reading the lecture rather than listening to it. In the same way that a lecturer might give you some reading to do, the study units contain instructions on when to read your set of books or other materials and practice some practical questions. Just as a lecturer might give you an in-class exercise or quiz, your study units provide exercises for you to do at appropriate point in time. Each of the study units follows a common format. The first item is an introduction to the subject matter of the unit and how a particular unit is integrated with the other units and the course as a whole. Followed by this is a set of objectives. These objectives let you know what you should be able to do at the end of each unit. These objectives are meant to guide you and assess your understanding of each unit. When you have finished the units, you must go back and check whether you have achieved the objectives. If you cultivate the habit of doing this, you will improve your chances of passing the course. The main body of the unit guides you through the required reading from other sources. This will usually be either from your set books or from your course guides. The following is a practical strategy for working through the course. Always remember that your tutor's job is to help you. When you need his assistance, do not hesitate to call and ask your tutor to provide it. Follow the under-listed pieces of advice carefully:-

1) Study this Course Guide thoroughly; it is your foremost assignment.
2) Organize a Study Schedule: refer to the course overview for more details. Note the time you are expected to suspend on each unit and how the assignments relate to the units.
3) Having created your personal study schedule, ensure you adhere strictly to it. The major reason that students fail is their inability to work along with their study schedule and thereby getting behind with their course work. If you have difficulties in working along with your schedule, it is important you let your tutor know.
4) Assemble the study material. Information about what you need for a unit is given in the 'overview' at the beginning of each unit. You will almost always need both the study unit you are working on and one of your set books on your desk at the same time.
5) Work through the unit. The content of the unit itself has been arranged to provide a sequence you will follow. As you work through the unit, you will be instructed to read sections from your set books or other articles. Use the unit to guide your reading.
6) Review the objectives for each unit to be informed that you have achieved them. If you feel uncertain about any of the objectives, review the study material or consult your tutor.

7) When you are sure that you have achieved the objectives of a unit, you can then start on the next unit. Proceed unit by unit through the course and try to space your study so that you keep yourself on schedule.
8) When you have submitted an assignment to your tutor for marking, do not wait for its return before starting on the next unit. Keep to your schedule. You are strongly advised to consult your tutor as soon as possible if you have any challenges or questions.
9) After completing the last unit, review the course and prepare yourself for the final examination. Check that you have achieved the objectives of the units (listed at the beginning of each unit) and the course objective (listed in this Course Guide).
10) Keep in touch with your study centre. Up-to-date course information will be constantly made available for you there.

Facilitators/Tutor and Tutorials

There are ten hours of tutorials provided in support of this course. You will be notified of the dates, times and location of these tutorials, together with the name and phone number of your tutor; as soon as you are allocated a tutorial group. Your tutor will grade and comment on your assignments, keep a close watch on your progress and on any difficulties you might encounter and provide assistance to you during the course. You must mail your tutor-marked assignment to your tutor well before the due date (atleast two working days are required). They will be marked by your tutor and returned to you as soon as possible.

Do not hesitate to contact your tutor by telephone, e-mail or personal discussions if you need help. The following might be circumstances in which you would find help necessary. Contact you tutor if:

i. You do not understand any part of the study unit or the assigned readers;
ii. You have difficulty/difficulties with the self assessment exercises;
iii. You have a question or problem with an assignment with your tutor's comments on any assignment or with the grading of an assignment.

You are advised to ensure that you attend tutorials regularly. This is the only opportunity to have a face to face contact with your tutor and ask questions. You can raise any problem encountered in the course of study. To gain the maximum benefit from course tutorials, prepare a question list before attending them and ensure you participate maximally and actively.

Summary

This course guide gives an overview of what to expect in the course of this study. ECO 154: Introduction to Quantitative Method II introduces you to the basic principles of statistics. It examines elementary concepts of statistics, measures of central tendency, measures of dispersion and measures of location. Other aspects of statistics addressed include moments, skewness and kurtosis as well as simple probability distribution and price number analysis.

Attention is drawn to the use of simple and lucid statistical issues in solving day to day economic and business related problems. The use of simple instructional language has been adequately considered in preparing the course guide.

NATIONAL OPEN UNIVERSITY OF NIGERIA

INTRODUCTION TO QUANTITATIVE METHOD II

ECO 154

SCHOOL OF ARTS AND SOCIAL SCIENCES

COURSE MATERIAL

Course Developer:

MAKU, Olukayode Emmanuel
School of Arts and Social Sciences,
National Open University, Nigeria.
14/16 Ahmadu Bello Way,
Victoria Island, Lagos.

Course Editor:
Dr. IBRAHIM, Adeolu
Bakare

CONTENT **PAGE**

Module 1: Basic Introduction

Unit 1: Meaning and Relevance of Statistics

Unit 2: Types/Branches of Statistics

Unit 3: Basic Concepts in Statistics

Unit 4: Collection of Data

Unit 5: Organization of Data

UNIT 1

MEANING AND RELEVANCE OF STATISTICS

Table of Contents

1.0 INTRODUCTION

The aspect of decision-making that has to do with numerical information is known as ***Statistics***. The word "Statistics" is often used t mean any of the following: numerical information; a summary of numerical information; a discipline. Statistics is generally used to describe a collection of numerical data; for example, population data, health statistics and school enrolment data, among others. It is also used to refer to the summary of a collection of numerical data such as the total, the minimum, maximum, range and average of such data of interest. As a discipline, statistics is the scientific method of decision-making under uncertainty when numerical data and calculated risks are involved. Statistics presents facts in a definite, lucid and concise form so that the facts are readily available for making valid conclusions. Statistics equally synthesizes large mass of data into simple format so that they convey meaning to the reader. The field of study called "Statistics" is fast becoming relevant and essential in all aspects of life because for decision to be appropriately taken, resources judiciously utilized and plans efficiency executed, data has to be collected, organised, analysed and interpreted. These are the bedrock of Statistics.

2.0 OBJECTIVES

At the end of his unit, you should be able to: -

i. Define the term "statistics".
ii. Outline the steps involved statistical enquiry
iii. Outline and discuss the relevance of statistics to other fields of study.
iv. Identify the limitation of statistics as a tool of decision making.

3.0 MAIN CONTENT

3.1 Meaning and Scope of Statistics

Statistics may be defined as the techniques by which data are collected, organised, summarised, presented and analyzed for the purpose of making reasonably valid decisions and especially under conditions of uncertainty. It is the science of making decisions under uncertainty, that is, making the best decision on the basis of incomplete information. Hence, in order to make a decision, the selection must be done without bias (random selection).

In summary, statistics is the science of collecting, organising, summarising and analysing information in order to draw conclusion. Statistics is classified as a science because it makes use of the tools of science such as principles, laws, experimentation, observation, theories/models, conclusion, generalization etc. The above definition can be considered in three parts. The first part of the definition states that statistics is the collection of information. The second part refers to the organisation and summarization of information. Finally, the third part states that the information is analyzed in order to draw conclusions. The definition implies that the methods of statistics follow a process.

The process of statistics can be categorised into four stages:

(i) Identifying the research objectives;
(ii) Collection of information needed to achieve the objectives;
(iii) Organization and summarization of information (data); and
(iv) The drawing of conclusions from the raw or analyzed information.

Hence, statistics is a process of factual data collection and analysis of the data. It involves collection of numerical facts in a systematic way. Statistics also involves the careful analysis of the data collected in form of tables and the interpretation of such data. It involves the use of scientific method of collecting, organizing, summarizing, presenting, analyzing data as well as drawing conclusions so as to take reasonable decision concerning a given phenomenon.

SELF ASSESSMENT EXERCISE 1

Explain the term “statistics” and state its major process.

3.2 Steps in Statistical Inquiry

A ***statistical inquiry*** is an investigation carried out to collect statistical data which may be analyzed and presented in the form that will aid effective decision making. Statistics as a discipline, if adequately carried out (following the steps in statistical inquiry), has significant impact on the decision making in almost all the fields of study. For statistics to be able to achieve its goals, the following steps must be properly followed: -

(i) Problem and the objectives should be properly stated.
(ii) Samples should be properly selected without bias.
(iii) Questionnaires should be well laid-out.
(iv) Data should be collected effectively and efficiently.
(v) Data should be properly organized.
(vi) Analysis and interpretation of data must be properly carried out.
(vii) Outcomes/Results of the analyses should be properly presented.
(viii) The report of the inquiry must be presented using simple and illustrative languages such as tables, charts or graphs.

SELF ASSESSMENT EXERCISE 2

Carefully select a social science problem and identify the systematic statistical approach of solving the problem.

3.3 Uses/Relevance of Statistics/Statistical Information

Although statistics is a powerful tool for analyzing numerical data, its application is widely seen in all fields of human endeavour. For instance, we apply statistics to: -

(i) **Physical Sciences: -** It determines whether or not experimental results should be incorporated into the general body of knowledge.
(ii) **Biological and Medical Sciences: -** Statistics guides the researcher in determining which experimental findings are significant enough to demand further study, or be tested more to meet human needs.
(iii) **Social Sciences: -** The roles of statistics in the social sciences cannot be ignored especially in economics, business and finance. The behaviour of individuals and organizations can be monitored through numerical data.
(iv) **Engineering: -** The professional field of engineering employs statistics in planning establishment policies and strong standards. A civil engineer may use statistics to determine the properties of various materials and perform some durability test.
(v) **Education: -** In the course of teaching, evaluation and assessment, statistics is very essential to analyze performance in examinations. The school headmaster can use statistics to write the curriculum, school enrolments, teachers/staff requirements, staff strength etc.
(vi) **Meteorology: -** Statistical information is also used in meteorology i.e. the science of weather prediction. In fact, the application of statistical techniques is so wide spread and the influence of statistics in our lives and habits is so great that the importance of statistics cannot be over-emphasized.

Apart from the relevance of statistics to the few chosen disciplines (as explained above), the following is the summary of general relevance of statistics in everyday life: -

(i) For summarizing large mass of data into concise and meaningful form leading to a better understanding of condensed data.
(ii) Giving visuals impact on data especially when presented in diagram, tables or charts.
(iii) Enabling comparison to be made among various types of data.
(iv) Making conclusions from data generated in pure experimental, social and behaviour research.
(v) Enabling a business establishment to make accurate, reasonable and reliable policies based on statistical data.
(vi) Predicting future events in daily life and business.
(vii) For the formation as well as testing of hypothesis.
(viii) For budgetary planning.
(ix) Widely used in industrial and commercial dealings as well as in government establishments.
(x) It's knowledge enables one to understood relevant articles in scientific journal and books

In spite of the relevance of statistics to everyday activities, the field of study (statistics) is limited by the following: -

(i) Statistics data or result is only an approximation of the total and therefore not entirely accurate in some cases. This is because not all the population will be covered for any statistical investigation.
(ii) Statistics if not carefully used can establish wrong conclusion and therefore it should only be handled by experts. Where experts are inadequate, some form of training should be conducted for those that may be required to carry out statistical research.
(iii) Statistics deals only with aggregate of fact as no importance is attached to individual items.

SELF ASSESSMENT EXERCISE 3

Outline and discuss any five uses of statistics to day to day activities of individuals, firms and the government.

4.0 CONCLUSION

This unit being the first of the introductory has been able to expose you to the meaning and scope of statistics, steps involved in carrying out statistical inquiry as well as the relevance of statistics to different fields of study and the day to day activities.

5.0 SUMMARY

Statistics is generally defined as science of data which involves collecting, classifying, summarizing or organizing, analyzing and interpreting data (numerical information) to be able to arrive at a statistical conclusion. In the course of carrying out statistical enquiry, a number of steps must be systematically carried out in order to make the result reliable. If this is properly done, results obtained from statistical research become useful in drawing inferences and conclusions.

Statistics as a field of study is widely applicable to most all the fields of study with particular attention to physical science, social science, engineering, education, meteorology etc. In spite of this, statistics has some limitation as a tool of analysis in different fields of study.

6.0 TUTOR MARKED ASSIGNMENT

1. How do you explain “statistics” to a layman?
2. What are the steps involved in statistical inquiry?
3. Explain the relevance of statistics to any five named field of study.
4. What are the limitations of statistics as a tool of analysis?

7.0 REFERENCES

Frank, O. and Jones, R. (1993). Statistics, Pitman Publishing, London.

Levin, R. I. (1988). Statistics for Managers. Eastern Economy Edition, Prentice Hall of India Private Limited.

Loto, Margret A., et al (2008). Statistics Made Easy. Concept Publication Limited, Nigeria.

Neil, A. Weiss (2008). Introductory Statistics (8th edition). Pearson International Edition, United State of America.

UNIT 2

BRANCHES OF STATISTICS

Table of Contents

1.0 INTRODUCTION

'Statistics' means numerical description to most people. Often the data are selected from some larger set of data whose characteristics we wish to estimate. We call this selection process sampling. For example, you might collect the ages of a sample of customers in a video store to estimate the average age or the most frequent age group of the customers in the store. Then, we can use the estimate to target the store's advertisements to the appropriate age group. It should be noted that statistics involves two different processes: -

(i) Describing sets of data; and
(ii) Drawing conclusion (making estimates, decisions, prediction etc.) about the set of data on the basis of sampling.

So, the applications of statistics can be divided into two broad areas: descriptive statistics and inferential statistics.

2.0 OBJECTIVES

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

(i) Identify the various branches of statistics.
(ii) Explain the various branches of statistics.

3.0 MAIN CONTENT

3.1 Descriptive Statistics

In ***descriptive statistics***, the data collected describes the situation that existed at the point in time when the census was taken. It provides a step by step detail of data available and collected at any given period. The important characteristics of descriptive statistics is that population to be described is definite. Descriptive statistics is the branch of statistics which deals with classification of data through the drawing of histograms that correspond to the frequency distribution which result after the data representation of data by different types of graphs such as line graphs, bar charts, pictogram, computation of sample mean, median and modes. The computation of variances, means, absolute deviations, deciles, range, percentile etc. are also regarded to as descriptive statistics. The main purpose of descriptive statistics is to provide an overview of the information collected.

Basically, the component of statistical process that deals with the organisation and summarization of information is referred to as *Descriptive Statistic*. It describes the information collected through numerical measurements, charts, graphs and tables. Therefore, descriptive statistics utilizes numerical and graphical methods to look for patterns in a data set, to summarize the information revealed in a data, and to present the information in a convenient form. Hence, descriptive statistics presents information in a convenient form at usable and understandable form in words with little numerical inclusion in the description of the data.

SELF ASSESSMENT EXERCISE 1

(1) What are the branches of Statistics?
(2) What is "Descriptive Statistics"?
(3) Outline the various tools used in descriptive statistical analysis.

3.2 Inferential Statistics

Inferential or analytical statistics is the second important branch of statistics. Most often, samples are carefully selected from population. On the basis of the sample, we infer things or conclusion about the population. This inference about populations on the basis of the sample is known as ***Statistical Inference***. In other words, statistical inference is the use of samples to reach conclusions about the populations from which those samples have been drawn. Inferential statistics is mostly linked with probability theory and estimate outcomes of events. Therefore, statistics of inference especially has to do with the measurement of chance.

We usually start with setting up a hypothesis (guess) or a number of hypotheses specifying our assumptions or guesses to be validated or refuted. This is usually stated at the beginning of the study. For example, we can test the hypothesis that the members of PDP are conservative in respect or economic policies while the members of ANPP take ***literal*** approach to economic policies. These testable assumptions are made from selected samples of members of the two political parties and not from the entire membership (the population). The result obtained can therefore be used to generalize for the entire political party provided the sampling is unbiased and statistical inquiry is properly carried out. This is technically called statistical inferences.

Making inferences is a question of chance; however, there are methods available to us to determine whether the results we obtain from a statistical investigation could be attributed to a chance occurrence even if its opposite were generally true.

On the other hand, we could also measure the odds that the result of our investigation is false. This will place us in a position to make right conclusion on a particular social or political phenomenon. When we do this, there is to a certain measure the possibility of the truth or the possibility of the falsehood of the assumption.

Inferential statistics utilizes sample data to make estimations, predictions or other generalization about a larger set of data. It uses methods that generalize results obtained from a sample of a population and measure their reliability.

Inferential statistics can be divided into two, namely: -

(i) Deductive statistics
(ii) Inductive statistics

(i) **Deductive Statistics: -** This is the act of drawing inferences about a sample using our knowledge of the population. The process involves arguing from the general (population) to specific (the sample). It is deductive inference when probability of an asset within a population context is obtained from a prior knowledge of the parameter of the distribution.

(ii) **Inductive Statistics: -** This is the process of drawing inference about the population from the sample. It is arguing from the specific (sample) to the general (population). Reasons of cost, time factor, accuracy and other constraints may make a complete enumeration (census) of the population impossible. The alternative is the use of concepts in probability to draw a sample from the population, obtain the estimate in the population parameter and test statement (hypothesis) about the parameter.

SELF ASSESSMENT EXERCISE 2

(i) What do you understand by "Inferential Statistics"?
(ii) Distinguish clearly between inductive statistics and deductive statistics.

4.0 CONCLUSION

This unit has been able to explain the basic branches of statistics – descriptive statistic and inferential statistics. It has also exposed you to the various classifications of inferential statistics which are deductive and deductive statistics.

5.0 SUMMARY

Statistics as a branch of study has two main divisions, namely descriptive statistics and inferential statistics. The descriptive statistics has to do with the collection, representation, organization and analysis of data while inferential statistics is concerned with interpretation, drawing inferences and measuring the reliability of the inferences drawn. It therefore consists of drawing and measuring the reliability about a population based on information obtained from a sample of the population. This branch of statistics can be inductive (drawing of inference(s) about the population from the sample) or deductive (drawing inference(s) about a sample using our knowledge of population.

6.0 TUTOR MARKED ASSIGNMENT

1. With appropriate and illustrative examples, explain the following branches of statistics.

    (i) Deductive statistics
    (ii) Inferential statistics
    (iii) Descriptive statistics
    (iv) Inductive statistics

2. Classify the following to descriptive or inferential statistics.

    (i) Representation of data
    (ii) Measures of central tendency
    (iii) Test of hypothesis
    (iv) Making judgements (by using either the sample or the population).

7.0 REFERENCES

Loto, Magret A. Et al (2008). Statistics Made Easy. Concept Publications Limited.

McClave, J. T. and Sincich, Terry (2009). Statistics (9th Edition). Pearson International Edition, New York.

Micheal, Sullivan (2005). Fundamentals of Statistics. Upper Saddle River Publishers, New Jersey.

Neil, A. Weiss (2008). Introductory Statistics (8th Edition). Pearson International, New York.

UNIT 3

BASIC CONCEPTS OF STATISTICS

Table of Contents

1.0 INTRODUCTION

Statistics like many other fields of study has its peculiar terms and terminologies. These terms are often used to describe the peculiarity of events and actions in the art of collecting, organizing, analyzing and interpreting data. Common among these terms are data, array variables, sample, population, sampling, sampling error and sampling frame.

2.0 OBJECTIVES

At the end of this unit, you should be able to define the basic concepts in statistics.

3.0 MAIN CONTENT

3.1 Data and Array

Data refers to a set or collection of usable information (singular: datum; plural: data). It may be about animate or inanimate objects. Raw data are collected data that have not been organised numerically. An example is the set of heights of 100 male students obtained from an alphabetical listing of university records.

The Standard English Dictionary defines data as facts and figures from which conclusions can be drawn. In general, however, statistical data refers to numerical description of quantitative aspects of a situation. Data are useful in providing an informal understanding of situations, with an overriding view to better decision making. Data collection is an activity or group of activities aimed at getting information, facts and figures to satisfy given decision objectives. It is perhaps the most crucial of all the stage in data analyses and interpretations.

Statistics as a body of methods intended for the study of numerical data, its first step in any statistical inquiry is the collection of the relevant numerical data. The data may be of two broad types: primary and secondary.

A primary data is a type of data collected directly from the source. These data are collected by the enquirer, either on his own or through some agency set up for the purpose, directly from the field of enquiry. This type of data may be used with greater confidence because the enquirer will himself decide upon the coverage of the data and the definitions to be used and as such will have a measure of control on the reliability of the data. The use of questionnaire, direct observation, personal interviews etc. are examples of primary sources of data.

The data may have already been collected by other agency, organization or institution (private or public) and may exist either in published or unpublished form. The researcher’s job is then to simply access them for research purpose. Such data is called ***Secondary Data***. In making use of secondary data, the enquirer has to be particularly careful about the nature, source and reliability of the data.

Arrays: ***An array*** is an arrangement of raw numerical data in ascending or descending order of magnitude. The major difference between data and array is that data is information presented as obtained without arrangement in any order but array is a set of quantifiable information arranged in order o magnitude (either in ascending or descending order). For example, scores of 10 students is given as 10, 12, 8, 7, 5, 6, 12, 10, 3 and 12. This is simply data. If this score is arranged in ascending or descending order as follows: -

3, 5, 6, 7, 8, 10, 10, 12, 12, 12 or

12, 12, 12, 10, 10, 8, 7, 6, 5, 3; it is called an *Array*.

CLASS ASSESSMENT 1

1. Distinguish clearly between primary and secondary data.
2. What is an array? How is it different from data?

3.2 Variable, Sample and Population

3.2.1 Variables

A variable may be defined as a symbol such as X, Y, x, y, H, α, β etc. that can assume any of a prescribed set of values, called “domain” of a variable. If the variable can assume only one value, it is called a “constant”. A variable that can theoretically assume any value between two given value is said to be ***continuous variable*** while the contrary is known as ***discrete variable***.

In statistics, variable are better described as the characteristics of the individuals within the population. It is the characteristic that varies from one person or things to another. Variables can be categorized into two major variables namely:

(i) Quantitative variables (ii) Qualitative variables

Qualitative or categorical variables allow for classification of individuals based on some attributes or characteristics. They are non-numerically valued variables, for example, colour (c) in a rainbow is a variable that can be red, orange, yellow, green, blue, indigo and violet. However, it is possible to attach subjective values to the colours e.g. indicate red by 1, orange by 2, and so on. Another example of qualitative variable is Gender. This is because it allows a researcher to categorize the individuals as male or female. It should be noted that arithmetic operation cannot be performed on the attributes "male" and "female".

Quantitative variables on the other hand are numerically valued variable. They provide numerical measures of individuals. Arithmetic operation such as addition, subtraction, multiplication and division can be performed on them such as meaningful results interpretation is obtained. Temperature is a quantitative variable because it is numeric and operations such as addition, subtraction etc. provides meaningful results. For example, 70^0C is 10^0C warmer than 60^0C. Other examples of quantitative variables are scores of students in an examination, ages of students, population of countries around the world, inflation rates, volume of sales etc. Quantitative variables can be classified into 2, namely: -

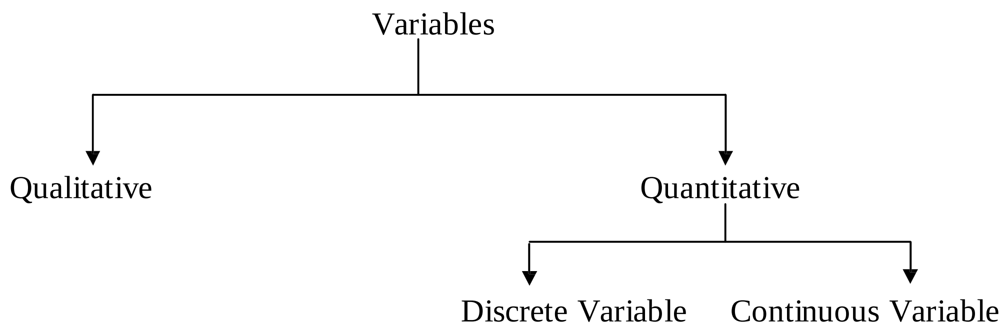
(i) Discrete variable
(ii) Continuous variable

A discrete variable is a variable whose possible values can be listed, even though the list may continue indefinitely. It usually involves a count of something, such as the number of sibling s a person has, the number of cars owned by a family, or the number of students in a statistics class. It does not allow any value between two given whole number i.e. fractions and decimals are not permitted. For instance, you cannot have 3½ siblings nor have 10.7 students in your class. Simply put, discrete variables are quantitative variables whose possible values can be listed. It is a quantitative variable that has either a finite number of possible or countable values. The term "countable" means the result from counting of positive/negative integers such as 0, 1, 2, 3...

Continuous variable on the other hand is a quantitative variable whose possible values form some intervals of numbers. Decimals, fractions and proportions are allowed e.g. temperature of cities, scores in examination (e.g. 67½ %), length of rubber (e.g. 10.5 mtrs) etc. A continuous variable is a quantitative variable that has an infinite numbers of possible values that are not countable.

It should be noted that the list of observation a variable assumes is called data. While gender is a variable, the observations, male or female are data. Qualitative data are observation corresponding to quantitative variable. Quantitative data are observations corresponding to discrete variable while continuous data are observations corresponding to a continuous variable.

The figure below illustrates the relationship among qualitative, discrete and continuous variables.



3.2.2 Samples

A sample is just part selected to represent the population. It is the collection of some members of the population based on a distinct definition. Covering each and every member of the population in the course of an inquiry, as stated earlier, is called complete enumeration. Sample are better used for research relative to complete enumeration (population) because it is carried out with less efforts, less cost, shorter time, greater accuracy, greater scope coverage and greater application.

A sample is simply put as a subset of the unit of a population. For example, instead of polling all 60 million registered voters in Nigeria during a presidential election year, a pollster might select and question a sample of just 1,500 voters. The 1,500 voters selected is the sample while the 60 million people is the population. If the sample of 1,500 voters is carefully selected, the results obtained for the sale can generalize for the population. A statistical inference or inductive statistics is an estimate, prediction or some other generalization about a population based on information contained in a sample. Hence, instead of examining the entire group, called the population or universe, one examines a small part of the group, called a sample which is that part of the population from which information is obtained.

3.2.3 Population

In statistics, the term ***"Population"*** refers to the whole of any group of individuals or items whose members (units) possess the same basic and clearly defined characteristics. For example, the population of a country includes every human being, adults, children, able and disabled people living within the boundary of a country; the population of books in the library refers to all books in the library and so on. We can have population of students, lawyers, teachers, doctors, vehicles, drivers, etc. in a given place at a given point in time.

Population is a collection of all possible usable information as may be required or as clearly defined. Population is relative. For example, a collection of all 200 level students of accounting is a population within the framework of 20 level accounting but a sample when it comes to the university population or students in the faculty of in which Accounting belongs. Population data is also known as ***census*** data.

SELF ASSESSMENT EXERCISE 2

1. Define the following terms as used in statistics
   (a) Variables
   (b) Samples
   (c) Population
   (d) Discrete variable
2. Distinguish clearly between
   (a) Discrete and continuous variable
   (b) Quantitative and qualitative variables
   (c) Sample and population

3.3 Sampling, Sampling Error and Sampling Frame

3.3.1 Sampling

A sample has been defined as the part of a population, carefully selected such that each member of the population is likely to be selected and what is obtained (sample) form a true representation of the population. The process of selecting samples from the population to ensure accuracy representation and unbiasedness is known as ***Sampling* or *Sampling Technique***. The goal of sampling is to obtain individuals (sample) that will participate in a study so that accurate information about the population can be obtained.

Sampling can either be random or non random. A random sample is one in which every unit of the population has the same chance of being selected. Thus, there should be no bias in the selection of any unit of the population. For example, in selecting a class representative from a group of 40 students, every student must have an equal chance of being selected as the representative for us to have a random sampling. There should be no conscious or unconscious bias in selecting a boy or girl, tall or short, fat or slim, brilliant or dull. Non random sampling is a method of collecting information where the interviewer selects his/her units/respondents not in accordance with the rule of chance. He/she is free; subject to some restriction to select his/her own units or respondents.

Such sampling is bias. It is approximate when the cost or efforts of getting data is enormous and there is no sample frame.

Some of the basic technique used in statistical sampling include: -

(i) Simple Random Sampling
(ii) Systematic sampling
(iii) Stratified sampling
(iv) Multi-stage sampling
(v) Quota sampling
(vi) Cluster sampling
(vii) Convenience sampling

(i) Simple Random Sampling: - A random sample of n experiment has been defined as the individual selected from population in such a way that every different sample of size n has an equal chance of selection. The word 'random' does not mean haphazard. It refers to a definite method of selection. A random sample is one in which every member of the population has equal chance of being selected in the sample. A technique for obtaining a random sample is to assign numbers or names to each of member of the population. Write these numbers on small pieces of paper; place them in a box and after mixing thoroughly, draw from the box in the lottery fashion. Although, random sampling is a long and expensive operation, it gives a reliable, unbiased picture of the whole population. This method of selection is possible where the population is homogeneous and relatively small; the sampling frame has to be completed. The most basic sample survey design is simple random sampling, which is often abbreviated as random sampling. Simple random sampling can be with replacement or without replacement. The sample obtained from simple random sampling is the simple random sampling.

(ii) Systematic Sampling: - For practical work, it is easier to select every earned item in a list of the population. This method is termed "systematic sampling" i.e. the first of the sample unit being selected by some random process. For instance, if the list comprises a population of say 25,000 and the sample required is 500 the selection of every 50th item i.e. $\frac{25,000}{500}$ will yield the required sample. Systematic (random) sampling is easier to execute than simple random sampling and usually provides comparable results. The sample obtained from systematic sampling is systematic sample. A systematic sample is obtained by selecting every Kth individual from the population. The first individual selected is a random number between 1 and K. The method is simple, save times and cheap. However, it is only applicable when the population is relatively small. The method is sometimes termed ***quasi-random*** based on the nature of the selection. For this method of sampling, there must be a complete and up-dated sampling frame. If adequate care is taken, its approximates is sufficiently close to simple random sampling. The method is not completely random because once the initial starting point has been determined, it follows that the remainder of the item selected for the sample are represented by constant interval.

(iii) **Stratified Sampling: -** When population is homogenous (same), simple random sampling or systematic sampling is appropriate. However, when population is heterogeneous i.e. comprises of different categories or form; a stratified sample can be taken. This is because the item/observation in different units will behave differently. In stratified sampling, the population is divided into strata, groups or blocks of units in such a way that each group is as homogeneous as possible (has same characteristics). Using stratified sampling technique, more accurate information is obtained though at a higher cost especially at the planning stage. A member of congress wishes to determine her constituency's opinion regarding estate taxes. She divides her constituency into three income classes: low income, middle income and upper income classes. She then takes a random sample of households from each income class. This is a practical stratified sampling.

(iv) **Multi-Stage Sampling: -** This is where a series of samples are taken at successive stages. For instance, in a case of national sample, the 1st stage will be to break the sample into the main geographical areas. In the 2nd stage, a limited number of town and rural districts in each of the states will be selected. In 3rd stage within the selected towns and rural districts, a sample of respondents allocated to each state is drawn. This may also involve the list in which certain households are selected and many more stages may be added. This process may be repeated until ultimately a number of quite small areas in different parts of the country have been selected. A random sample of the relevant people within each of these groups/areas is then chosen for the research.

(v) **Quota Sampling: -** To economist and business managers, time and cost is taken into consideration in sample data, for this reason, a method of sampling known as quota sampling is extensively employed by many organisations. The interviewer is interested in carrying out number of interview with samples that conforms to certain requirement or law or rule. This rule may be restrictions imposed by sex, age, social group, geographical location or state of origin. A very good example is the selection of political officers (e.g. Senate President) from a particular geographical location. An important advantage of quota sampling is that it is not expensive in terms of money and time to conduct. Like other non-random sampling in quota sampling, there is no basis for computing the standard error of statistic and therefore the result cannot be taken precise.

(vi) **Cluster Sampling: -** In his technique, the country is divided into small areas; similar to multi-level sampling method. The interviewers are sent to the person they can find in an area of interest. Cluster sampling involves the use of the reference map of the area of interest. With the aid of the map, the area to be surveyed may be divided into smaller units and random sampling will be used to select some of the areas. The group of individuals so formed is known as

cluster. The most important advantage of cluster sampling is that, it is not expensive to conduct compared with other method of sample survey. However, the basic problem that may hamper the application of this method is that the individual cluster may be heterogeneous; therefore, the final solution of 'cluster' will involve a random sampling.

(vii) Convenience Sampling: - Convenience sampling is probably the easiest sampling method. A convenient sample (obtained from convenience sampling) is a sample in which the individuals are easily obtained. There are many types of convenience samples, but probably the most populace are those in which the individuals in the samples are self selected (i.e. the individual themselves decide to participate in a survey or not). Examples of self-selected sampling include phone in polling where a radio personality will ask his or her listeners to phone the station to submit their opinions. Another example is the use of internet to conduct surveys. It is closely related to Haphazard selection in which the interviewer makes a 'random' selection according to the dictate of his mind. This method encourages bias into the selection of the units of interest.

3.3.2 Sampling Frame

For a well representative sample to be selected, there is need for a detailed account of all units of the population from which a sample is to be selected. Such a detailed account can be found in the sampling frame. A sampling frame is the list of all the population units from which sample units are identified and selected. The sampling frame must be complete and constantly revised and updated. Examples of sampling frame include a school register, a voters' list, and the telephone directory.

3.3.3 Sampling Error

Sampling error is the error that results from using sampling to estimate information regarding a population. This type of error occurs because a sample gives incomplete information about the population. It measures the deviation between the sample's behaviour and the population's characteristics. Sampling errors may be as a result of: -

(i) Error due to bias.
(ii) The nature of the questionnaire.
(iii) Memory error (the respondent may give wrong information when the event being investigated has a long time).
(iv) Coding error (the use of wrong codes in carrying out the statistical survey and collation).
(v) Editing error (error that emerges in coding e.g. writing 1997 instead of 1977).
(vi) Error due to tabulation (sometimes errors emerge as a result of wrong tabulation of statistical information).

(vii) Error in the sharing of questionnaires.

3.4 Advantages and Disadvantages of Sampling

Advantages: - The merits of using sampling include:

(i) Sampling saves time and money.
(ii) Smaller number involved in sampling make “call backs” possible and allow for thorough check on returned questionnaire.
(iii) Compared with the process of complete enumeration, less number of people will be employed and effectively used, in the course of sampling.
(iv) The result from sampling is obtained within a short period of time.
(v) The error can be measured and readily handled.
(vi) The result obtained from sampling may be more accurate if well conducted.

Disadvantages: - The demerits of using sampling include:

(i) It cannot be used where the sampling frame is adequate or completely not available.
(ii) Sampling method breeds sampling errors which if not properly handled, may affect the result.
(iii) It can be manipulated to suit the purpose of investigator or collector.

4.0 CONCLUSION

The unit has been able to expose you to the nitty-gritty of the essential concepts, terms and terminologies used in statistics.

5.0 SUMMARY

Data are sets of statistical information while array is a set of data arranged in ascending or descending order of magnitude. Variables are the instruments used to obtain data. Variables may be quantitative, qualitative, discrete or continuous. Sample is the subset of the entire enumeration (census) upon which statistical investigations are carried out. A good sample is expected to give exact behaviour or result as the population. The techniques of obtaining sample are called sampling. Sampling may be random (when each item has equal chance of being selected) or non-random (when there is bias in the selection process). Sampling techniques are numerous. Some of them include simple random sampling, stratified sampling, systematic sampling, cluster sampling, multi-stage sampling, convenience sampling etc. Sampling has a lot of advantage but it is not devoid of some demerits.

6.0 TUTOR MARKED ASSIGNMENT

1. Write short but explanatory notes on the following sampling techniques

   (a) Convenience sampling
   (b) Multi-stage sampling
   (c) Systematic sampling
   (d) Simple random sampling
   (e) Stratified sampling

2. Identify the type of sampling used in addressing the following statistical inquiry.

   (a) In an effort to determine consumer satisfaction, United Airlines randomly select 50 flights during a certain week and survey all passengers on flight.

   (b) In an effort to identify whether an advertising campaign has been effective, a marketing firm conducts a nationwide poll by randomly selecting individuals from a list of known users of the product.

   (c) A radio station asks its listeners to call in their opinion regarding the use of American forces in peacekeeping mission.

   (d) A farmer divides his orchard into 50 subsections, randomly selects 4 and samples all of these within the 4 subsections in order to approximate the yield of his orchard.

   (e) A survey regarding download time on a certain website is administered on the internet by a market research firm to anyone who would like to take it.

3. A manufacturing company has 80 employees and each employee keep a record of his or her production on a weekly basis. Find below a record of production for 80 employees for the month ended September 30th.

40	22	50	61	30	58	51	75
58	70	73	59	49	55	63	38
87	57	23	41	60	57	52	77
37	62	53	83	48	73	28	31
21	76	32	57	53	25	42	63
95	54	64	39	82	54	33	45
48	22	53	65	26	65	87	43
51	66	34	78	55	44	27	74
89	46	67	45	30	57	97	81
43	28	99	47	79	56	68	35

Required: Form an array from the set of the raw data.

7.0 REFERENCES/FURTHER READINGS

Frank, O. and Jones, R. (1993). Statistics, Pitman Publishing, London.

Levin, R. I. (1988). Statistics for Managers. Eastern Economy Edition, Prentice Hall of India Private Limited.

Loto, Margret A., et al (2008). Statistics Concept Publication Limited, Nigeria.

McClave, J. T. and Sincich, Terry (2009). Statistics (9th Edition). Pearson International Edition, New York.

Micheal, Sullivan (2005). Fundamentals of Statistics. Upper Saddle River Publishers, New Jersey.

Neil, A. Weiss (2008). Introductory Statistics (8th edition). Pearson International Edition, United State of America.

UNIT 4

COLLECTION OF DATA

Table of Contents

1.0 INTRODUCTION

In the last unit, you were taught the meaning and various types of data. In this unit, we shall be discussing the various ways of collecting data as well as the challenges inherent in collecting data.

2.0 OBJECTIVES

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

(i) Identify and explain the various ways of collecting data; and
(ii) Explain the problems associated with data collection.

3.0 MAIN CONTENT

3.1 Methods of Collecting Data

Whether the data to be collected is primary or secondary, the collection may be done in either of the two ways: -

(a) Complete enumeration; and
(b) Representative enumeration

In ***complete enumeration***, each and every individual of the group to which the data relates is covered and information gathered for each individual separately. In ***representative enumeration***, only a representative part of the group is covered, either because the group is too large or because the number of items on which information is sought is too large. This is the case of sampling.

By data collection, we mean the principles, techniques, approaches and strategies used to seek for information especially for statistical analyses. Business data are collected in normal cause of administration and not specifically for statistical purpose. Proper care must however be taken when collecting data or either statistical or non-statistical use. Once a researcher has decided the type of data to employ, the next thing is to set out strategies on how the data is collated. Generally, there are numerous sources of data, the commonest ones are: -

(i) Direct observation
(ii) Personal interview
(iii) Use of questionnaire
(iv) Reports/Results of experiment
(v) Extraction from already established results

(i) Direct Observation: - This is the easiest of all and all it requires is to observe all the items in a specified population and draw conclusion from them. For example, a child psychologist might observe and record the level of aggressive behaviour of a sample of nursery students playing on a school's playground, a secondary school student may observe the most bought car by the member of staff in his school by observing the brand of cars parked at the parking space. The researcher or data collector makes no attempt to control any aspect of the experimental units, he only observes.

This method may entails sending observers to record what actually happens while it is happening at the current period. It can either be participatory or non participatory. The method is cheap and easy to understand. It equally has limited chance of being inconsistent. However, it cannot be used for collection of a vast range of data; it takes time and may be ambiguous when dealing with voluminous data. Direct observation as a method reduces the chance of incorrect data being recorded but it is limited by the size of observation. Thus, the method may involve observation, measurement or counting.

(ii) Personal Interview: - This is the method of collection which involves more than one person. There are two categories of people involved; the ***interviewer*** and the ***interviewee***. The interviewer carefully request response from the questions, he asks orally from the interviewee. The (interviewer) researcher asks the interviewee some questions verbally to provide a guide to what is being investigated. The interviewee may be expected to record facts and information as accurately as possible. This may be done electrically or otherwise. Interview is equally used in personal investigations and team investigation (where investigators go to the interviewee in group), it is sometimes called ***delegated personal investigation***. Personal interview is sometimes called ***Survey***. Survey requires researcher to sample a group of people, ask one or more questions and record the responses. Probably, the most familiar types of survey is the public opinion poll (e.g. political poll). Survey can be conducted through mail, with telephone interviewer or with in-person interviews. Although, in-person interview are more expensive than the mail or telephone surveys, they may be necessary when complex information is to be collected.

Interview provides first hand information, allows for a feed back and has high probability of receiving accurate information. However, it may be bias due to the personality, mood and pretence of the interviewee on issues.

(iii) **Questionnaire: -** A questionnaire is a carefully designed form to be completed by the respondent (the person that fills the form). The questionnaire may seek to know the bio-data (age, sex, marital status, state of origin, nationalities etc.); or contain direct questions on the main issue to be investigated. The researcher (the person that prepares the questionnaire for research purpose) prepares a set of question to be responded to by respondent in written form. The questions are made clear, detailed and unambiguous. After the respondents must have properly responded to the questionnaires in writing the researcher then arrange the questionnaire for collation and analysis. The list of the questions may be sent by post or by mail and may be returned via the same medium.

This method is cheap (especially if the questionnaire is posted or mailed and received the same way). It also gives the researcher and the respondent enough time to gather, respond to the questionnaire and analyse it. Results of the use of questionnaire are usually of high reliability because of the confidentiality attached to it. The disadvantage of questionnaire includes false data from respondents, misunderstanding of the question, forgetfulness etc.

Questionnaire method of data collection is cheap. You don't have to go to the respondent, it could be sent through electronic mail or be posted. It could equally allow the respondent to supply the information required by the researcher. However, there could be some wastage of time in responding to the questionnaire if the respondent is not responsive enough. This may cause some delay in the conduct of the research.

(iv) **Reports / Results of Experiment: -** This method is of interest to the production managers, engineers the scientist etc. It requires carrying out experiments (not necessarily in the laboratory) and using the result to determine the behaviour of certain things or circumstance. The researcher exerts strict control over the units (people, objects or things) in the study. For example, a medical study which intends to investigate the potency of aspirin in preventing heart attacks. Such experiments may be carried out on a number of patients that use the tablet (aspirin) and whatever result obtained is subject to generalization based on its validity and reliability. Results must be properly recorded after experiments have been carefully carried out to meet the standards required. However, the major disadvantage of this method is that, it is rather expensive and may also give a misleading result if the experiment is not carefully carried out.

(v) Extraction from already Established Result: - This is one of the commonest ways of collecting secondary data. In this approach, the data of interest has already been collected for the researcher and available in published source, such as book, journals, newspaper, the internet etc. Users of data extracted from already established results are not the original collector of the data. As a result of this, he may not have thorough understanding of the background of the data, compared to the original data collector. As such, the user of already established data may be ignorant of the limitations and assumptions taken in compiling the data. However, it is a fast way of collecting data because the data is already in existence only to be accessed and utilized.

CLASS ASSESSMENT EXERCISE 1

1. Outline the advantages and disadvantages of collecting statistical data via each of the followings means
   - (i) Questionnaire
   - (ii) Direct observation
   - (iii) Report or Result of Experiment
2. What are the limitations of obtaining statistical data or information from an already established result?

3.2 Problems of Data Collection

Data collection can be difficult or inaccurate sometimes. The absence and unavailability of accurate statistical data may be due to all or some of the following reasons:

- (i) Lack of proper communication between users and producers of statistical data.
- (ii) Difficulty in estimating variables which are of interest to researchers and planners.
- (iii) Ignorance and illiteracy of the respondents.
- (iv) High proportion of non response due to suspicious on the part of the respondent.
- (v) Lack of proper framework for which samples can be selected.
- (vi) The wrong ordering of priorities including misdirection of emphasis and bad utilization of human and material resources.

CLASS ASSESSMENT EXERCISE 2

What are the problems you may likely encounter as a student of statistics in finding out the factors responsible for poor performance of secondary school students in Mathematics?

4.0 CONCLUSION

This unit has been able to introduce you to the various ways through which researchers obtain information and data. Data is indispensable tool in statistical analyses. This call for serious attention to ensure that data used in analysis are free of errors, bias and other problems. The units also highlight the likely problems researchers and statisticians encounter in the course of collecting data.

5.0 SUMMARY

Data collection may involve the direct participation of the collectors as we observe in primary data collection (direct observation, use of interview and questionnaire as well as carrying out experiments). In case of secondary data, e.g. extracting already established result, the data collector is not involved in the collation of the data. The data already exists, he only access it and make proper use of it. Each of primary and secondary sources of data collection has their own shortcomings.

6.0 TUTOR MARKED ASSIGNMENT

1. Outline and discuss any four ways of collecting data.
2. Compare and contract direct observation and questionnaires as primary sources of data.
3. Choose a research of your interest; enumerate how you will collect the data and the problems you are likely to encounter in the process of collecting the data.

7.0 REFERENCES/FURTHER READINGS

Levin, R. T. (1988). Statistics for Managers; Eastern Economy Edition. Prentice - Hall of India Private Limited.

Magaret, A. Loto, et al (2008). Statistics Made Essay. Concept Publication Limited, Lagos, Nigeria.

Mc Clave, James T. and Sincich, Terry (2009). Statistics (11th Edition); Pearson International Edition, Upper Saddle River, United State of America.

UNIT 5

ORGANIZATION OF DATA

Table of Contents	
1.0	Introduction
2.0	Objectives
3.0	Main Content
3.1	Organization of Ungrouped Data
3.2	Organization of Grouped Data
4.0	Conclusion
5.0	Summary
6.0	Tutor Marked Assignment
7.0	References/Further Reading

1.0 INTRODUCTION

In the last unit, you have been exposed to the various approaches of collecting data. Data collection is not the only activity in statistic. Collected data has to be properly arranged and organised. This will make analysis and interpretation easier.

This unit shall discuss the techniques and terminologies used in arranging and organizing data such that the information or data collected becomes compact and easy to manage. The unit shall equally introduce you to how data are organized for grouped data as well as for the ungrouped data.

2.0 OBJECTIVES

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

(i) Distinguish clearly between grouped data and the ungrouped data.
(ii) Explain the terminologies used in organizing both grouped and ungrouped data.
(iii) Organize raw data in either grouped or ungrouped pattern i.e. construct frequency distribution tables.

3.0 MAIN CONTENT

3.1 Organization of Ungrouped Data

An ***ungrouped data*** is an array of information such that each item has its own individual frequency or occurrence. The information or data are not in group but treated as individuals. In order to prepare an ungrouped frequency distribution, each data is counted physically and recorded with the use of strokes (tally), which is later written out in figures under the frequency column. This in totality is called *frequency distribution*. A frequency distribution lists the number of occurrences for each category of a data. An ungrouped frequency distribution shows at glance the number of times each of the data occurs (frequency) and the sum of times all data occur ($\sum$ **f**). The method is stronger and more compact than array presentation.

Example: Given the ages of students in a class as: 10, 8, 9, 7, 8, 10, 11, 8, 12, 6, 7, 8, 10, 8, 7, 7, 9, 9, 8, 10, 8, 8, 8, 7, 8, 11, 11, 8, 10, and 11.

Required:

(a) Present the raw data in an array using
  (i) Ascending order of magnitude
  (ii) Descending order of magnitude

(b) Present the data using ungrouped frequency distribution table

(c) From the table, find;
  (i) The most occurring students age
  (ii) The difference between the age of the oldest student and the age of the youngest student.
  (iii) How many students are older than a 7-year old student?

Solution

(a) Array in ascending order of magnitude (smallest to highest): 6, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 9, 9, 9, 10, 10, 10, 10, 10, 11, 11, 11, 11, 12.

Descending order of magnitude (largest to smallest): 12, 11, 11, 11, 11, 10, 10, 10, 10, 10, 9, 9, 9, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 7, 7, 7, 7, 7, 6.

(b)

Ages	Tally	Frequency
6	l	1
7	~~llll~~	5
8	~~llll~~ ~~llll~~ l	11
9	lll	3
10	~~llll~~	5
11	llll	4
12	l	1
		$\sum$ **f** = 30

Note: $\sum$ **f** = total of the frequency column i.e. (1 + 5 + 11 + 3 + 5 + 4 + 1).

(c) (i) The most occurring age is 8 years old with 11 students.
  (ii) The oldest of student = 12 years

The youngest student = 6 years

$\therefore$ Difference between the age of the oldest and the youngest student = 12 – 6 = 6 years.

(iii) The students older than a 7-year old student are those above the age of 7 i.e. those whose age 8 years and above.

Years	Frequency
8	11
9	3
10	5
11	4
12	1
	$\sum \mathbf{f} = 24$

Therefore, 24 students out of the 30 students are older than a 7-year old student.

CLASS ASSESSMENT EXERCISE 1

The places of birth of some past United States Presidents are given:

Virginia, Massachusetts, Virginia, Virginia, Virginia, Massachusetts, South Carolina, New York, Virginia, Virginia, North Carolina, Virginia, New York, New Hampshire, Pennsylvania, Kentucky, North Carolina, Ohio, Ohio, Ohio, Vermont, New Jersey, Ohio, New Jersey, Ohio, New York, Ohio, Virginia, Ohio, Vermont, Iowa, New York, Missouri, Texas, Massachusetts, Texas, California, Nebraska, Georgia, Illinois, Massachusetts, Arkansas and Connecticut.

Required:

(a) Construct a frequency distribution of the state of birth.
(b) Use the table to answer the following questions:
  (i) Which of the states has produced the president most often?
  (ii) In which states has the president being produced only once?
  (iii) Which states produce the presidents twice?
  (iv) How many states have produced the president for more than once?

3.2 Organization of Grouped Data

Data organized and summarized in such a way that information are classified by groups or classes, are called grouped data. The tabular arrangement of data by classes together with the corresponding class frequencies is called frequency distribution or frequency table of grouped data. Although, the grouping process generally destroys much of the original detail of the data, an important advantages is gained in the clear “overall’ picture that is obtained and in the vital relationships that are thereby made evident. There are some terminologies used in preparing grouped data. They are:

(i) Class/Groups
(ii) Class size/class width
(iii) Class interval and class limits
(iv) Frequency
(v) Cumulative frequency and cumulative frequency distribution
(vi) Relative frequency and relative frequency distribution.
(vii) Class boundary
(viii) Lower class boundary/lower class point.
(ix) Upper class boundary/upper cut point.
(x) Class mark or the midpoint.

(i) Class/Group: - This is the categories for grouped data e.g. 21 – 30, 31 – 40, 41 – 50, 51 – 60 etc.

(ii) Class Size/Width: - This is the number of items (data) that form a group. In the example above (21 – 30, 31 – 40, 41 – 50, 51 – 60 ...), the class size is 10 because there are 10 item in each group. Closely related to this is the class limits e.g. 31 – 21 = 10, 40 – 30 =10. The class size is also called the class length, denoted by *c*. It could also be obtained by finding the difference between the two consecutive lower or two consecutive upper class boundaries.

(iii) Class Intervals and Class Limits: - A symbol defining a class such as 41 – 50, is called the class interval. The end numbers 41 and 50 are called class limits, the smaller (41) is the lower class limit and the larger (50 is the upper class limit. The term class and class intervals are used interchangeably although the class interval is actually a symbol for the class. A class interval (atleast theoretically) has either no upper limit or no lower class limit indicated is called an open class interval. For example, referring to age group of individuals, the class interval “65 years” and above is an open class interval.

(iv) Frequency: - This is the number of observations that fall in a class.

(v) Cummulative Frequency and Cummulative Frequency Distribution: - Cummulative frequency is the summing up of the frequencies of each class while cummulative frequency distribution is the listing of the classes and their cummulative frequency.

(vi) Relative Frequency and Relative Frequency Distribution: - The relative frequency is the proportion or percent of observation within a category/group/class relative to the entire sample size. Simply put, it is the ratio of the frequency of a class to the total number of observations. The relative frequency distribution on the other hand, is the listing of all classes and their relative frequencies. It lists the elative frequency of each category of data.

$$\text{RelatiR ffReqR ntC} = \frac{\text{frequency of a grouped/unit}}{\text{um of a frequencie}}$$

(vii) Class Boundaries: - In an attempt to introduce continuous variable into class intervals, class boundaries emerge. For example, if heights are recorded to nearest inch, the class interval 60 – 62 theoretically includes all measurement from 59.5 to 62.5. These numbers, indicated briefly by the exact numbers 59.5 and 62.5 are called class boundaries, or true class limits. Hence, the class boundary is 59.5 – 62.5.

(viii) Lower Class Boundaries/Lower Cut Point: - This is the least component of a class boundary for each group. In the example above, the lower class boundaries is 59.5. It is the smallest value that could go into a class.

(ix) Upper Class Boundaries/Upper Cut Point: - This is the highest opponent of a class boundary for each group or class. In the example above, the upper class boundary is 62.5. Upper class boundary or upper cut point is the highest value that could go into a class. It is equally the smallest value that could go into the next higher class.

(x) Class Mark / Midpoint: - The class mark is the midpoint of the class interval and is obtained by adding the lower and the upper class limits or those of the class boundaries and divide by 2. Thus, the class mark or midpoint of $60 - 62$ ts $\frac{60+62}{2}$ of $\frac{59.5+62.5}{2} = 61$. For the purpose of further mathematical analysis, all observations belonging to a given class interval are assumed to coincide with the class mark.

Example: Given the days to maturity for 40 short term investments as:

70	64	99	55	64	89	87	65
62	38	67	70	60	69	78	39
75	56	71	51	99	68	95	86
57	53	47	50	55	81	80	98
51	36	63	66	85	79	83	70

Required:

(a) Prepare the frequency distribution for the data using the class interval 31 – 40, 41 – 50, 51 – 60 …, along with the class boundaries.

(b) In addition to (a) above, prepare the columns for the midpoint, relative frequency, cummulative frequency and the cummulative relative frequency.

Solution

Class Groups	Class Boundaries	Tally	Frequency	Midpoint	Cumm. Freq.	Rel. Freq.	Cumm. Rel. Freq
31 – 40	30.5 – 40.5	lll	3	35.5	3	0.075	0.075
41 – 50	40.5 – 50.5	ll	2	45.5	5	0.050	0.125
51 – 60	50.5 – 60.5	~~llll~~ lll	8	55.5	13	0.200	0.325
61 – 70	60.5 – 70.5	~~llll~~ ~~llll~~ ll	12	65.5	25	0.300	0.525
71 – 80	70.5 – 80.5	~~llll~~	5	75.5	30	0.125	0.750
81 – 90	80.5 – 90.5	~~llll~~ l	6	85.5	36	0.150	0.900
91 – 100	90.5 – 100.5	llll	4	95.5	40	0.100	1.000
			$\sum \mathbf{f} = 40$				

CLASS ASSESSMENT EXERCISE 2

The marks below are obtained from an examination conducted for 64 students: -

78	62	44	46	34	53	40	51	67	55	46
40	50	51	48	56	81	61	47	40	46	49
35	50	56	48	58	70	38	50	36	61	67
59	64	58	44	54	55	76	40	64	66	46
60	62	70	40	54	36	48	54	31	52	30
50	40	50	56	36	77	71	73	66		

Required:

(a) With the class size of 8, starting with 30 –, how many classes will the scores be grouped.
(b) Prepare a frequency distribution table for the score including the class boundaries, the frequency columns.

4.0 CONCLUSION

The unit has been able to expose you to the techniques involved in arranging or organizing data (either using the grouped or ungrouped approach). You have also been properly introduced to some terminologies used in preparing frequency tables of grouped data. These terminologies shall be useful for you in subsequent modules.

5.0 SUMMARY

After data has been collected, there is a need to arrange and organize them into a compact form. This is usually done with a table known as the frequency distribution table or the frequency table. Such tables can be prepared for both grouped and ungrouped data.

6.0 TUTOR MARKED ASSIGNMENT

1. a) Arrange the numbers: 17, 45, 38, 27, 6, 48, 11, 57, 34 and 22 in an array.
   b) Determine the difference between the highest and the least observation.
2. The final grades in mathematics of 80 students in a state university examination are recorded in the accompanying table.

68	84	75	82	68	90	62	88	76	93
73	79	88	73	60	93	71	59	85	75
61	65	75	87	74	62	95	78	63	72
66	78	82	75	94	77	69	74	68	60
96	78	89	61	75	95	60	79	83	71
79	62	67	97	78	85	76	65	71	75
65	80	73	57	88	78	62	76	53	74
86	67	73	81	72	63	76	75	85	77

(a) From the above, find the highest grade, the least grade and the difference between the highest and the least grade.

(b) Find the grades of the
  (i) Five highest ranking students.
  (ii) Five lowest ranking students.
  (iii) The student ranking 10th highest.
  (iv) Number of students who received grades of 75 or higher.
  (v) Number of students who received grades below 85.
  (vi) Percentage of students who received grades higher than 65 but not higher than 85.

(c) Prepare the frequency distribution table, using the class interval of 50 – 54, 55 – 59, 60 – 64, and so on.

(d) Obtain the midpoint, cummulative frequency, class boundaries, relative frequency and the cummulative relative frequency for each group.

7.0 REFERENCE/FURTHER READING

Frank, O. and Jones, R. (1993). Statistics; Pitman Publishing Limited, London.

Spiegel, M. R. and Stephen, L. J. (2004). Statistics; Shaum Outline Series. Tata McGraw-Hills limited, New Delhi.

Sullivan, Michael (2005). Fundamentals of Statistics; Anointed Instructor's Edition. Pearson Prentice Hall, Upper Saddle River, New Jersey.

Module 2: Representation o Data

Unit 1: Tables

Unit 2: Graphs

Unit 3: Charts

Unit 4: Histogram and Curves

UNIT 1

TABLES

Table of Contents	
1.0	Introduction
2.0	Objectives
3.0	Main Content
3.1	Definition and Scope of a Table
3.2	Features/Properties of a Good Table
3.3	Importance of Tables
4.0	Conclusion
5.0	Summary
6.0	Tutor Marked Assignment
7.0	References/Further Reading

1.0 INTRODUCTION

In the last module, statistics has been defined as the process of collecting, organizing, analyzing and interpreting data in order to make appropriate decision. This implies that statistics is not limited to collection of data; the data collected has to be property organized and represented.

In statistical analyses, information gathered can be presented in the form of table (e.g. descriptive tables, frequency distribution table etc), charts, graphs and curves. Table is perhaps the most widely acceptable way of representing data because of the ease in interpreting information presented on it. It is widely used in all disciplines i.e. it is not limited to statistics, Mathematics, Economics, Social Sciences and Allied Sciences. The presentation of data in the tabular form gives an idea of the distribution of the information gathered for further evaluation.

2.0 OBJECTIVES

At the end of this unit, you should be able to;

(i) Define the word "Table" and states its components;
(ii) Explain the features or properties of a Good Table and
(iii) Outline the importance of Table in Statistic and other related disciplines.

3.0 MAIN CONTENT

3.1 Definition and Scope of a Table

A ***table*** is an orderly arranged list of information, facts or data. It is usually set out in rows and column in an attempt to summarize large quantity of data by presenting it in a condensed, compact and lucid manner. Table can also be described as a methodological and orderly arrangement of information using rows and columns. The vertical classifications or divisions form the column while the horizontal classification from the row. A table is simply an orderly arrangement of information showing the relationship between or among variables.

The presentation of data in form of tables is preferred by statisticians because a table may be so constructed as to include the whole max of data in a concise form. A table should have a number of parts:

(i) **Title: -** A title gives a brief description of the content. Every table is expected to have a title.

(ii) **Stub: -** The extreme left part of a table which is meant to give a description of the rows is called the stub of the table.

(iii) **Caption: -** The upper part of the table which gives a description of the various columns is the caption of the table. The caption may have to be accompanied by a mention of the units of measurement for the data of each column and also by column numbers.

Note: Title, stub and capture, taken together, are said to form the ***box head of the table***.

(iv) **Body: -** The body is the principal part of the table, where the figures are exhibited.

(v) **Foot note: -** Most tables also have footnotes, where the sources of the data are indicated and explanations given regarding the scope, source notes and reliability of items are stated.

Examples of tables are illustrated below:

Exports and Imports of Oil seeds and Vegetable Oils in Nigeria, 1991 – 1992

	Quantities (Tonner)				Value (₦m)			
Exports	**Dec. 1991**	**Dec. 1992**	**June – Dec. 1991**	**June – Dec. 1992**	**Dec. 1991**	**Dec. 1992**	**June–Dec. 1991**	**June–Dec. 1992**
H.P.S Groundnut Kernel	3024	171	6262	8894	2.56	57.45	123	175
Castor oil	375	6538	9604	7137	10.21	171.14	398	196
Imports	**April 1991**	**April 1992**	**June – July 1991**	**June – July 1992**	**April 1991**	**April 1992**	**June – July 1991**	**June – July 1992**
Mustard Seeds	6	530	5554	2388	0.05	10.7	64	25
Soya bean oil	7961	4465	9006	7971	176.7	168	285	294

Source: Agricultural situation in Nigeria (Various Issues). Economic and Statistical Review (Various Issues) of the CBN

Example 2

Consider the following marks obtained by students in an entrance examination.

A's Score English: 40%, Maths: 60% General Knowledge: 80%
B's Score English: 80%, Maths: 60% General Knowledge: 60%
C's Score English: 80%, Maths: 40% General Knowledge: 60%
D's Score English: 60%, Maths: 50% General Knowledge: 80%

It is observed that this piece of information does not make for easy comprehension. You are required to:

(i) Show the distribution of the students' performance in the courses with the use of a hypothetical table.
(ii) Compare and contrast the array of the information given and the table constructed

Solution

Performance of students in an Entrance Examination

Student	Math (%)	English (%)	General Knowledge (%)	Total Marks
A	60	40	80	180
B	60	80	60	200
C	40	80	60	180
D	50	60	80	190
Total Score	***210***	***260***	***280***	***750***

Form the table, we can see at a glance the relative performance of the four students, we can also interpret the relative performance in the 3 subjects. Based on the table, we can draw some inferences or conclusions. The aggregate marks in the last column shows that "student B" has the highest total mark of 200. On subject basis, student A and B are the best in Maths, B and C the best in English and Student A and D, the best in General Knowledge.

By and large, the table provides us with detailed and more analytical information than the array of the scores.

SELF ASSESSMENT EXERCISE 1

Consider the following information on the performance of some students in the post UME examination.

Matta: English 57%, Math 50%, Current Affairs 81%
Idoko: English 87%, Math 79%, Current Affairs 67%
David: English 69%, Math 62%, Current Affairs 61%

(a) Present the performance of the students in a table

(b) Interpret your findings.

3.2 Features or Properties of a Good Table

Although, tables are very useful in statistical analyses, they serve their purposes more effectively and efficiently if their construction adhered to some basic principles, laws and tenets. A good table should satisfy the following properties:

(i) It must have a neat outlay and be easily understood i.e. self contained and self explanatory.

(ii) It must have a general explanatory title or heading. The title should be clear, unambiguous and concise. It must indicate the purpose of the table.
(iii) The units of measurement must be clearly defined and shown.
(iv) It must contain foot notes and source notes to describe the details and the origin of the table.
(v) It must have column title to indicate the type of items classified in the column.
(vi) It must have row title to indicate the type of item classified in the row.

SELF ASSESSMENT EXERCISE 2

Outline the any fire properties of a good Table as a medium of data representation.

3.3 Importance of Table

The specific importance of Table in statistics and other relates disciplined are:

(i) It is used to interpret data more vividly and more clearly
(ii) Data in table can be used for comparative analysis
(iii) Quick decisions can be taken based on information derived from tables
(iv) Information from tables occupies less space
(v) Tables reveal at a glance, the information conveyed on the data
(vi) The data presented in a table can be used to forecast future performance of the variables involved.
(vii) A clearer relationship between or among variables is shown with the use of tables.
(viii) Summary of information being presented are shown with the use of tables.
(ix) Some tables present information which are useful for further research
(x) Required information or figures are easily located from tables.

CLASS ASSESSMENT EXERCISE 3

Of what relevance are tables to statistical analyses?

4.0 CONCLUSION

As one of the major way of representing data, table has been defined as an orderly presentation of data in raw and column from which statistical inferences can be drawn. The unit also outlines the constituents, properties and the importance of tables in statistics and every day activities.

5.0 SUMMARY

A statistical table is an orderly presentation of data in rows and columns. It has the primary advantage of condensing and thereby facilitating comparison of data collected. It also allows for easy summarization of item and detection of error. Repetition of explanations, figures headings and notes can be avoided through the contraction and the use of statistical tables. A statistical table is equally easier to comprehend than a max of unorganized data (array). However, a good table has to be properly presented and adhere to all the principles and rules guiding the construction of tables in order to be able to achieve its set down goals.

6.0 TUTOR-MARKED ASSIGNMENT

(1) In 1976, 578,000 candidates under 13 years of age took entrance examination to government secondary school. Out of these candidates, 335,000 of them were boys and the rest girls. When compared with 1975, the total number increased by 3,100, the increase of 7200 boys was nearly offset by a decrease in the number of girls. By 1976, 300,000 candidates of 11 years of age (110,000 girls) sat for the entrance examination. Compared with 1975 figure, the total increased by 8,200 out of which there were 4,900 boys, 32,000 girls and 57, 000 boys aged 10 sat for the entrance examination in 1975 and in 1976, the figure were increased by 5,000 girls and 8100 boys.

Construct a table form the above information and show all relevant total and sub-totals comment on the results.

(2) The table below gives the average annual output of major food crops in Nigeria.

	Average annual output of Major food crops in Nigeria (1960 - 1971) Years (output in million Tones)					
Crops	**1960-62**	**1963-65**	**1966-1968**	**1969-71**	**Total Output 1960-1971**	**Average Output (1960-1971)**
Sorghum	3.979	4.172	3.126	4.041		
Millet	2.576	2.64	2.273	2.957		
Rice	0.220	0.215	0.312	0.277		
Maize	1.101	1.148	1.095	1.425		
Cowpea	0.481	0.622	0.593	0.930		
Yam	12.901	14.818	11.649	11.997		
Cassava	7.212	7.982	6.175	6.175		

Source: Agric Development in Nigeria (Federal Ministry of Agric and Rural Development), Rural Economic Survey of Nigeria (Federal Ministry of Statistics).

Required

(a) Compute the column for total output (1960 – 71) and the average output (1960 – 71).

(b) Which of the crop contributed least to food crop between 1966 and 1968 and by what quantity?

(c) What proportion of the total output is yam between 1963 and 1965?

(d) Calculate the percentage of the average output of 1969 – 1971 to the total average output of maize from 1960 – 1971.

7.0 REFERENCES/FURTHER READING

Frank, O. and Jones, R. (1993). Statistics; Pitman Publishers Limited.

Nwabuoku, P. O. (1986). Fundamental of Statistics; Roruna Book, Enugu, Nigeria.

UNIT 2

GRAPHS

Table of Contents	
1.0	Introduction
2.0	Objectives
3.0	Main Content
3.1	Meaning and features of Graphs
3.2	Importance of Graphs in Statistics
4.0	Conclusion
5.0	Summary
6.0	Tutor Marked Assignment
7.0	References/Further Reading

1.0 INTRODUCTION

In the last unit we discussed table as a way of presenting data. Table is not only tool used for data presentation, graphs, chart and curves are also used. Curves, charts and graphs are closely related. They are used to give pictogram presentation of collected data. They have visual appeal to readers by showing relationship between two or more variables and show trends hence they facilitate ready comparison. Glancing at diagrams, charts and graphs, readers are attracted to read the accompanying report in order to know the causes of variations in the phenomena depicted by them. Diagrams, charts and graphs have more visual appeal to those who cannot appreciate the value of figures. They can be used to show the growth or otherwise of variable e.g. fluctuations in prices of goods performance of foreign trade, etc.

However, diagrams, charts and graphs have their shortcomings. They many over-simplify the information they depict. They may be too sensational as to distort the actual observation recorded. They may become absurd when drawn from incomplete or inadequate data. More so, over dependency on the use of diagrams, charts and graphs may not allow the users to pay attention to the root cause of the problems they reveal.

2.0 OBJECTIVES

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

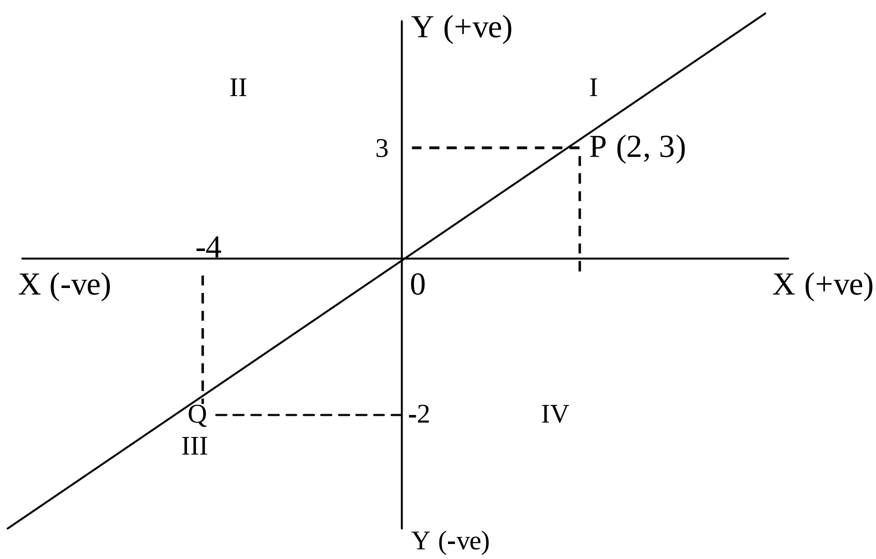
(i) Define and explain the features of Graphs;
(ii) Outline the importance of Graphs; and
(iii) Construct Line graphs, given some information.

3.0 MAIN CONTENT

3.1 Meaning and Features of Graphs

A graph is a pictorial presentation of the relationship between variables. It can also be defined as a functional relationship between two variables. A graph may be depicted by lines to establish relationship between variables. Although, graphs are curves are often used interchangeably, they are technically different from one another. Graphs are mostly depicted with straight lines or joined points (curves) while curves establish relationship with the use of free-hand sketch only and not with the use of lines. Therefore all curves are graph but not all graphs are curves.

Graphs are usually drawn on a co-ordinate plane called XY plane which is divided into four regions denoted by I, II, III, and IV as shown below. Each of the region is called a quadrant.



Point O is called the origin, or Zero point. The value of X and Y at the points where the perpendicular meet these axes are called rectangular coordinates; or simply coordinates, of P, denoted by (X, Y). The coordinate of X is sometimes called abscissa and Y is called the ordinate of the point. From the above diagram P is on (2, 3). Therefore the abscissa of point P is 2 and the ordinate is 3. Two points can be joined together with a straight line resulting into a Graph. For instance P and Q form line PQ, where PQ is a straight line curve/graph.

Graphs are expected to have the following features:

(1) It must have a clear title.
(2) It must be properly are carefully labelled with the units of each axis.
(3) It must be properly scaled so as to give an accurate impression of the presentation.

(4) It must be able to forecast and estimate values.

SELF ASSESSMENT EXERCISE 1

1. What are Graphs?
2. What are the relevance of graphs, charts and diagrams in representation of data.
3. Explain the following terms as applied to graphical layout:
   (i) Ordinate
   (ii) Coordinate
   (iii) Abscissa
   (iv) Origin
   (v) Rectangular Coordinate
4. What are the essential features of statistical graphs?

3.2 Importance of Graphs in Statistics

Like tables, graphs equally have relevance to the study of statistics and other related disciplines. Some of the importance of graphs in statistics includes:

(i) Graphs give instantaneous impression about the information being presented.
(ii) Functional relationship expressed in graphs show a dearer picture of the relationship between variables.
(iii) Graphs are illustrative and descriptive tool for economic analyses.
(iv) Data presented in table are better understood when transposed into graphs.
(v) Graphs are important for quick estimation and forecasting purposed.

CLASS ASSESSMENT EXERCISE 2

What are the relevance of Graphs to statistics?

3.3 Linear Graphs

Although, graphs may be formed through the joining of points on a plane co-ordinate, the shape of graph may be curves or straight line. Linear relationships among variables are usually depicted with straight lines. Such lines may be downward sloping or upward sloping. Non linear relationship are shown with different forms of curves

Later part of this module shall treat curves and charts in detail. This unit shall therefore forms on linear graphs. A linear graph is a straight line graph that shows the relationship between two variables, one on the X–axis and the other on the Y axis (see coordinate plane). The vertical axis (Y) represents the dependent variable while the X axis which is the horizontal axis, represents the independent variable. Therefore a straight line graph has the general functional equation stated as $Y = \mathbf{1} \pm bX$.

Where a = Intercept of the graph and b is the slope.

Example 1

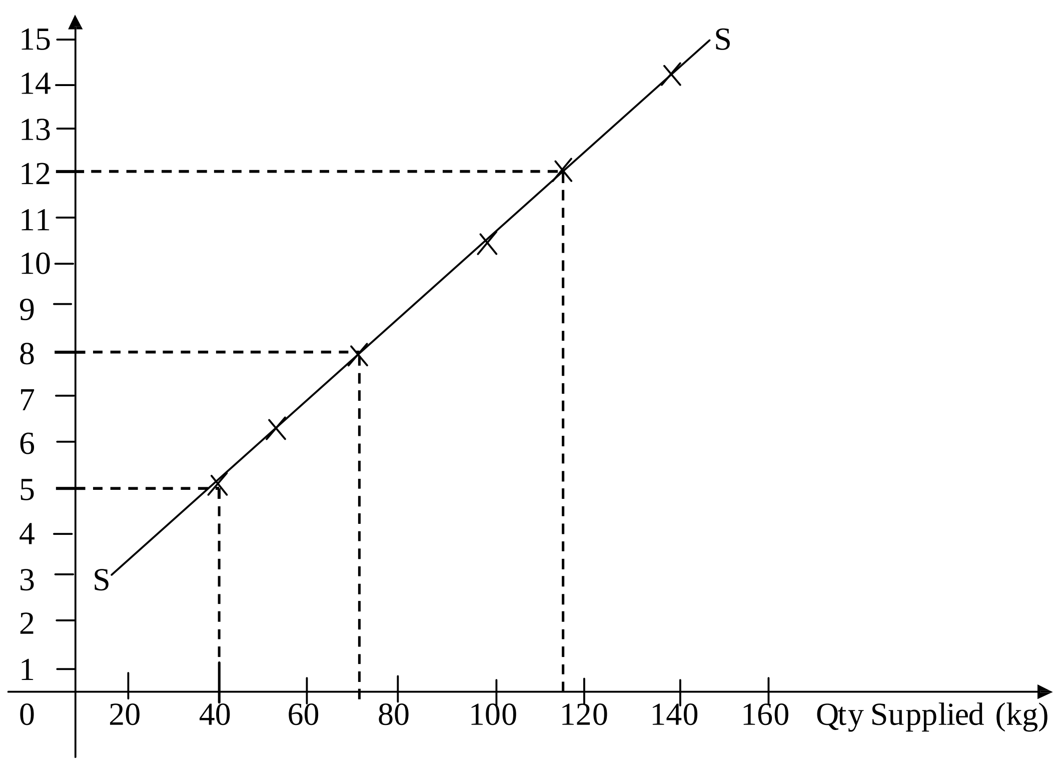
The table below shows the relationship between the quantity supplied and price of a good.

Price (₦)	5	7	9	11	13	15
Quantity Supplied (Kg)	40	60	80	100	120	140

Required: (a) Present the information using graph placing price on the vertical axis and the quantity supplied on the horizontal axis.

(b) From the graph, obtain the quantity supplied when the price is ₦12.

(c) At what price is 70 kg supplied?



(b) At price ₦12, quantity, the supplied is 110kg

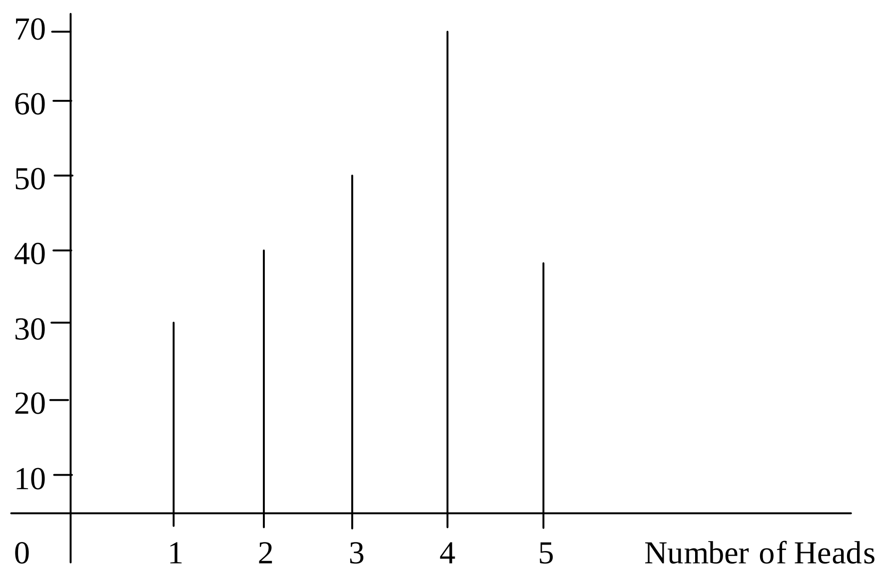
(c) 70kg of the good is supplied when price is ₦8

(2) Given the data in the table below

No of Heads	No of tosses (frequency)
0	30
1	40
2	60
3	50
4	70
5	40
6	290

Required: Present the data using rod graph

Solution



SELF ASSESSMENT EXERCISE 3

1. Given that quality demand (Q_d) function of a production is given as $Q_d = 200 - \frac{\ }{2}p$

   (a) Prepare a table for the quantity demanded given that the prices are ₦10, ₦20, ₦30 … ₦100
   (b) Use the table obtained in (a) above to graph the quantity demand
   (c) From the graph find: -
      (i) The quantity demanded when the price is ₦85.
      (ii) At what price is quantity demand equals 170 units?

2. Given the age distribution (in %) for a country as

Ages	0–10	11– 20	21– 30	31– 40	41– 50	51– 60	61– 70	71– 80	> 80
%	5	10	12	15	18	5	10	10	15

Present the data using rod graphs.

4.0 CONCLUSION

In this unit, the meaning and properties of graph have been examined. Equally examined is the importance of graph, construction of different forms of line graph as well as estimating values from the graph.

5.0 SUMMARY

Graphs are functional relationship between two variables (dependent and independent). Usually, graphs are drawn on coordinate planes and it could either take the form of straight line or hand sketch curves. Graphs have numerous importances to statistical values and interpretations can be obtained from graphs.

6.0 TUTOR MARKED ASSIGNMENT

1. The consumption function of a consumer is given as C = 60 + 0.4 Yd. Where C = consumption expenditure and Yd is the disposable income.

(a) Complete the table below in ₦ ’000

C										
Yd	10	20	30	40	50	60	70	80	90	100

(b) Present the information using a straight line graph.
(c) From the graph drawn find
   (i) C when Yd = 72
   (ii) Yd when C = 100

2. The table below shows the scores of students in a physics test.

Scores	1	2	3	4	5	6	7	8	9	10
No of students	8	7	9	7	4	5	8	6	5	0

Required:

(a) Present the data using rod graph
(b) How many students took part in the test?
(c) What is the most frequent score in the test?
(d) Does you graph indicate the highest score?

3. Given the Nigeria's population figure between 1960 and 2006 as follows:

Census Year	1960	1970	1980	1990	2000	2004
Population (millions)	60	70	80	90	120	123

Present the information in a line diagram.

7.0 REFERENCES/ FURTHER READING

Spiegel. M. R and Stephen J. L (2000). Statistics (Third Edition). Schaum's Outline Sense Data McGraw –Hill Publishing Company Limited, Delhi.

UNIT 3

CHARTS

Table of Contents	

1.0 INTRODUCTION

Apart from tables and graph, statistics data are equally presented with the use of charts and curves. Various forms of charts and curves are used in statistics to demonstrate data presentation. Charts have a ready appeal to the eye and are therefore they are helpful in convey the significance of the data quickly. They can also indicate the trend of a time series, together with the nature of fluctuations, if there are any. They are equally useful in detecting mistakes in the data. Charts are often used together with tables and sometimes with textual statements. The commonest forms of charts are bar charts, pie chart and Z – charts and pictograms.

It should be noted that charts, and the other pictorial diagrams have their shortcomings or limitations. These include:

(i) They reveal only the general nature of the data.
(ii) They cannot show as much details as a table or even a textual statement will do.
(iii) Their construction requires much more time than tables, for the same set of data.

2.0 OBJECTIVES

At the end of the unit you should be able to:

(i) Define different forms of charts, curves and diagrams
(ii) Present data in bar charts of different forms as well as in pie charts and Z–Charts.

3.0 MAIN CONTENT

3.1 General Overview

Charts are diagrammatic representation of data with the use of bars, shapes, curves and other illustrative objects. Commonest among the forms of charts used in statistic are bar charts, pie charts, Z–Charts etc.

3.2 Bar Charts

It consists of bars of rectangle which are of equal width with each of its length corresponding to the frequency or quantity they are representing. The bars are separated from one another by equal intervals of gaps. Each axis of the diagram should be properly scaled with the indication of the unit(s) of measurement along each axis. Bar charts or bar diagrams is another common mode of data representation. They are more generally applicable than line diagrams in the sense that they may be used for senses varying either over time or over space. They could also be used to represent distribution of multi-dimensional variables e.g. multiple and component bar charts.

Bar charts are constructed by representing the categories (units/classes) of the qualitative variables by bars, where the height of each bar is the class frequency, class relative frequency, class relative percentage, or cumulative percentages. Each category of data is placed on the horizontal axis and the frequency or relative frequency or percentage cumulative frequency on the vertical axis. Bar charts (simple vertical bar charts) are like a histogram. However, to avoid confusion bar graphs/ chart and histogram, we position the bars in the bar charts, so that they do not touch each other.

CLASS ASSESSMENT EXERCISE 1

(1) What are Charts?
(2) What are the basic rules needed to be observed in constructing bar charts?

3.2 Forms of Bar Chart

Different forms of bar charts used in statistics include:

(i) Simple vertical Bar Chart
(ii) Simple Horizontal Bar Chart

(iii) Multiple Bar Charts
(iv) Percentage Multiple Bar Chart
(v) Component Bar Charts
(vi) Percentage component Bar Charts
(vii) Pareto Charts

3.2.1 Simple Vertical Bar Charts

This involves the drawing of bars upright or vertically. It is a chart in which the length of the bars indicated the magnitude of the data. Each of the vertical bars shows the magnitude of the occurrence of the situation under study. It is sometimes called ***vertical frequency bar chart***.

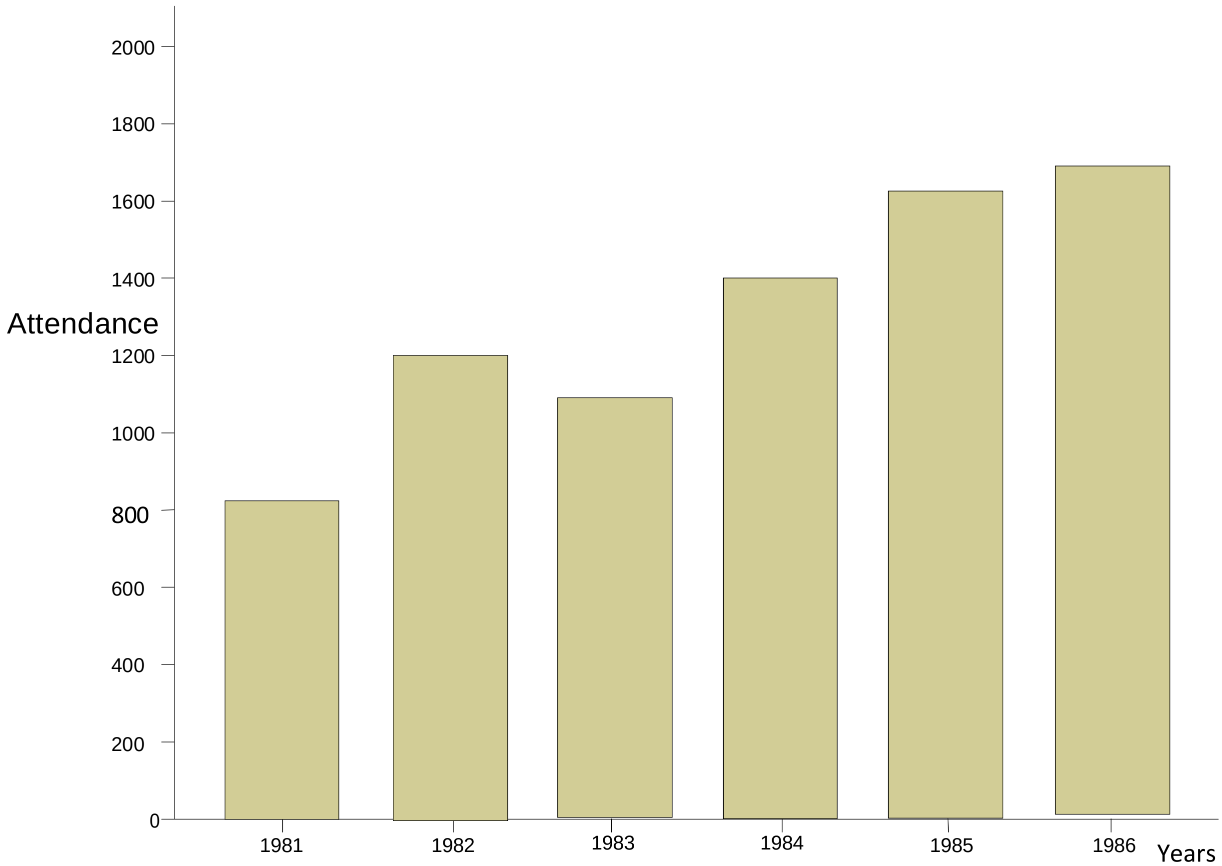
Example

Peace House has the following 6 years projection for those that will attend its annual teachers' conference. Present the date in a vertical bar chart.

Year	**Attendance**
1981	800
1982	1200
1983	1100
1984	1400
1985	1600
1986	1700

Solution

Simple vertical Bar chart showing Attendance in Teachers conferences (1981 - 1986)



3.2.2 Simple Horizontal Bar Chart:

The simple horizontal bar charts involve drawing bars horizontally thereby presenting the frequency on the horizontal axis and the variable in the vertical axis. It should be noted that, like vertical bar charts, the bars should be of equal with and the spaces between the bars should be uniform to show relative measurement. It is sometimes called horizontal frequency bar chart.

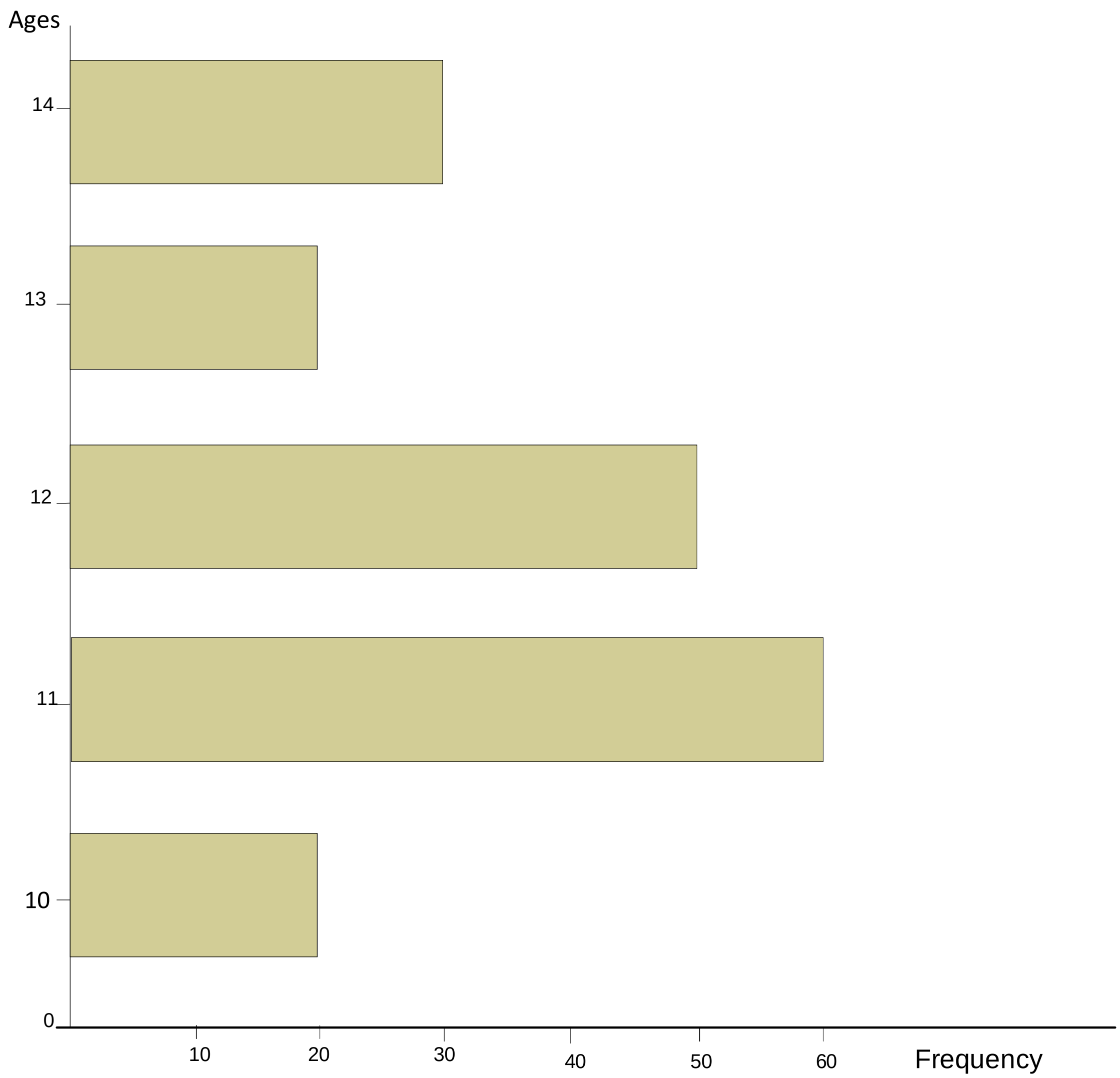
Example

Given the data below:

Ages	10	11	12	13	14
Frequency	20	60	50	20	30

Required: Represent the information in a simple horizontal bar chart.

Solution



3.2.3 Multiple Bar Charts

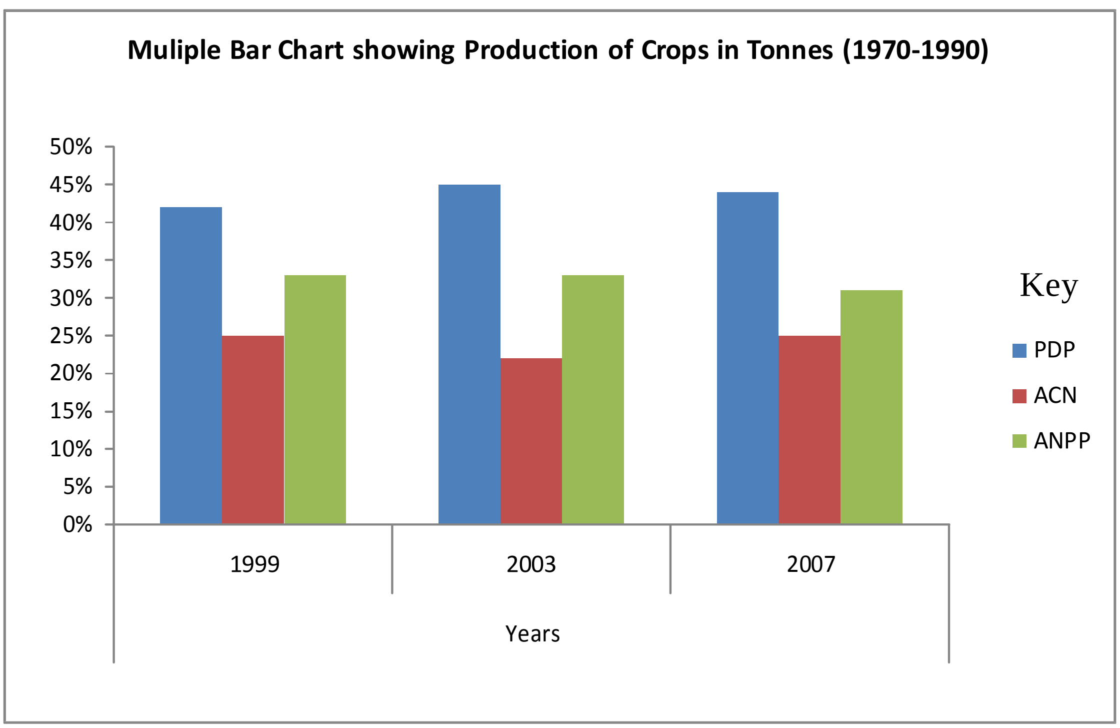
This consists of bars of two or more variables being represented. It is also called compound bar chart or side-by-side bar graph. The items represented are usually differentiated with different designs or colour. This type of chart helps us to compare two or more sets of data. Data sets should always be compared by using relative frequencies, because different sample or population sizes make comparison using frequencies difficult. Key is expected to show what each sign, colour or design represents.

Example

The production of rubber, cotton and coffee in 1970, 1980 and 1990 is given as follows:

Years	**Production in Tonnes**			
	Cocoa	Coffee	Rubber	Total
1970	10	15	35	60
1980	20	35	45	100
1990	40	90	100	230

Required: Present the information using Multiple Bar Chart.



3.2.4 Percentage Multiple Bar Chart

This mainly shows the relative values of the components expressed as the percentage of the total. Therefore, the sum of each component in percentage adds up to 100.

Example

Given the information below:

Political Parties	**Vote in thousand**			
	1999	**2003**	**2007**	**2011**
PDP	25	40	70	90
ACN	15	20	40	70
ANPP	20	30	50	40
TOTAL	60	90	160	200

Required: Present the information using a percentage Multiple Bar Chart.

Solution

Percentages

1999

PDP = $\frac{25}{60} \times \frac{00}{} = 42\%$

ACN = $\frac{5}{60} \times \frac{00}{} = 25\%$

ANPP = $\frac{20}{60} \times \frac{00}{} = 33\% \therefore 42\% + 25\% + 33\% = 100\%$

2003

PDP = $\frac{40}{90} \times \frac{00}{} = 45\%$

ACN = $\frac{20}{60} \times \frac{00}{} = 22\%$

ANPP = $\frac{30}{90} \times \frac{00}{} = 33\% \therefore 45\% + 22\% + 33\% = 100\%$

2007

PDP = $\frac{70}{60} \times \frac{00}{} = 44\%$

ACN = $\frac{40}{60} \times \frac{00}{} = 25\%$

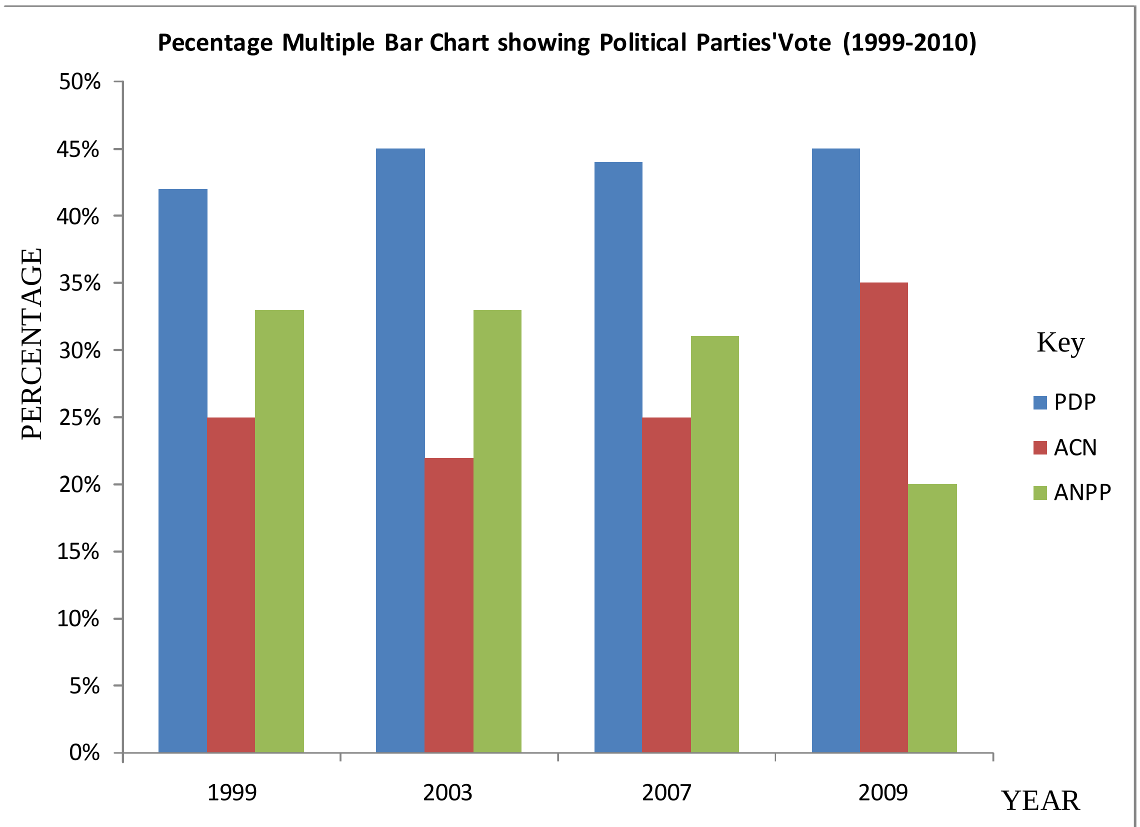
$\text{ANPP} = \frac{500}{60} \times \frac{00}{} = 31\% \therefore 44\% + 25\% + 31\% = 100\%$

2009

$\text{PDP} = \frac{90}{200} \times \frac{00}{} = 45\%$

$\text{ACN} = \frac{70}{200} \times \frac{00}{} = 35\%$

$\text{ANPP} = \frac{40}{200} \times \frac{00}{} = 20\% \therefore 45\% + 35\% + 20\% = 100\%$



3.2.5 Component Bar Chart

A component bar chart shows the breakdown of the total values for a given information into their component parts. It is meant to show the division of a whole into its constituent parts in bars. The bars are drawn on one another. A bar is therefore divided into different parts for the sake of easy comparison. Just like multiple bar chart, key is provided to define what each component of the bar stands for?

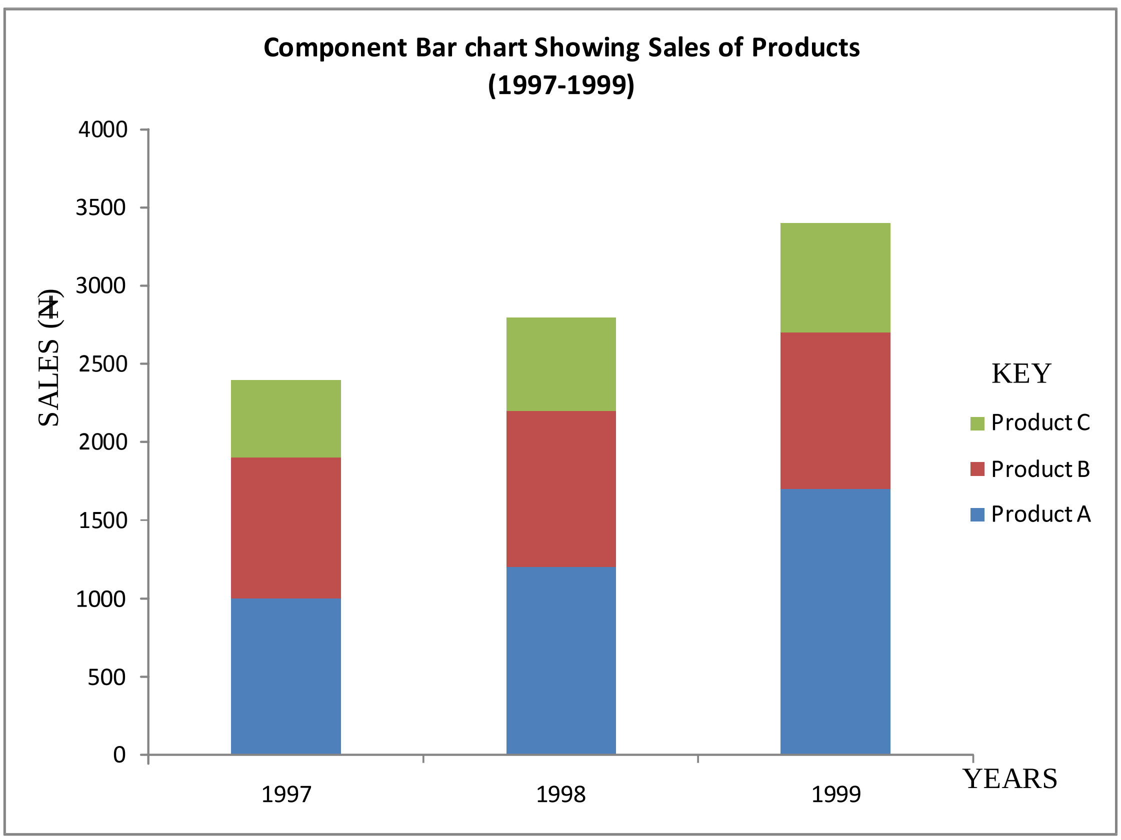
Example

Given the sales of 3 products in a market as follows:

Products	**Sales (₦)**		
	1997	1998	1999
Product A	1000	1200	1700
Product B	900	1000	1000
Product C	500	600	700
Total	2400	2800	3400

Required: Present the information using component bar chart

Solution



3.2.6 Percentage Component Bar Charts

This is a component bar chart in which each constituent part of the bar is presented as the percentage of the total. The constituent parts are placed on one another in cumulative percentages. Therefore, the percentage component of each bar amounts to 100%

Example

The table below shows the domiciles of students Attending University of Lagos 1972 – 76 (Hypothetical)

Year	**Domiciles in Nigeria**			
	Total students	West	East	North
1972	120	60	48	12
1973	150	90	30	30
1974	160	80	50	30
1975	200	110	50	40

Required: Present the data with the use of percentage component bar chart.

Solution

Percentages

1972: West = $\frac{60}{20} \times \frac{00}{} = 50\%$

East = $\frac{48}{20} \times \frac{00}{} = 40\%$

North = $\frac{2}{20} \times \frac{00}{} = 10\% \therefore 50\% + 40\% + 10\% = 100\%$

1973: West = $\frac{90}{50} \times \frac{00}{} = 50\%$

East = $\frac{30}{50} \times \frac{00}{} = 20\%$

North = $\frac{30}{50} \times \frac{00}{} = 20\% \therefore 60\% + 20\% + 20\% = 100\%$

1974: West = $\frac{80}{60} \times \frac{00}{} = 50\%$

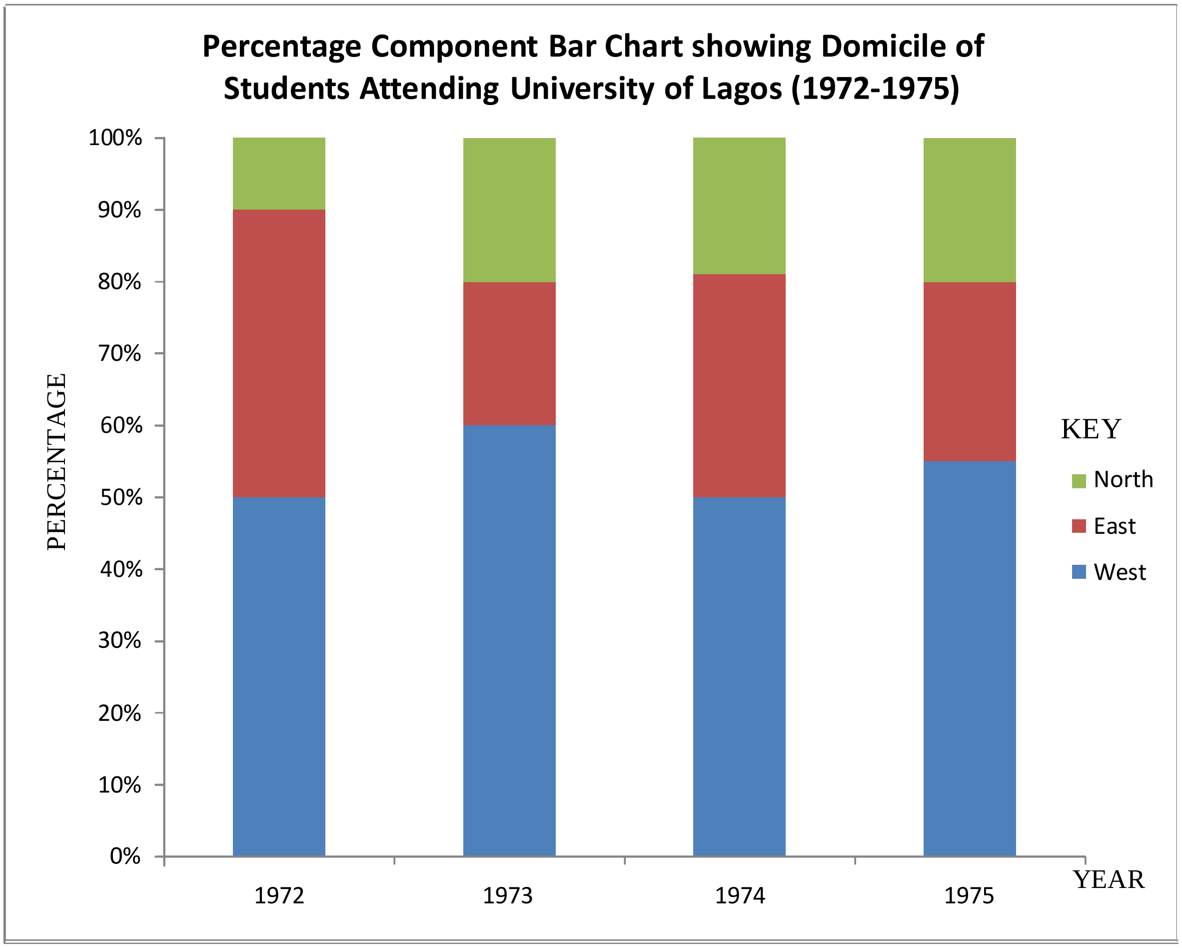
East = $\frac{60}{60} \times \frac{00}{} = 31\%$

North = $\frac{30}{60} \times \frac{00}{} = 19\% \therefore 50\% + 31\% + 19\% = 100\%$

1975: West = $\frac{0}{200} \times \frac{00}{} = 55\%$

East = $\frac{50}{200} \times \frac{00}{} = 25\%$

$$\text{North} = \frac{40}{200} \times \frac{00}{} = 20\% \therefore 55\% + 25\% + 20\% = 100\%$$



3.2.7 Pareto Chart

A Pareto chart is a bar graph whose bars are drawn in decreasing order of frequency or relative frequency.

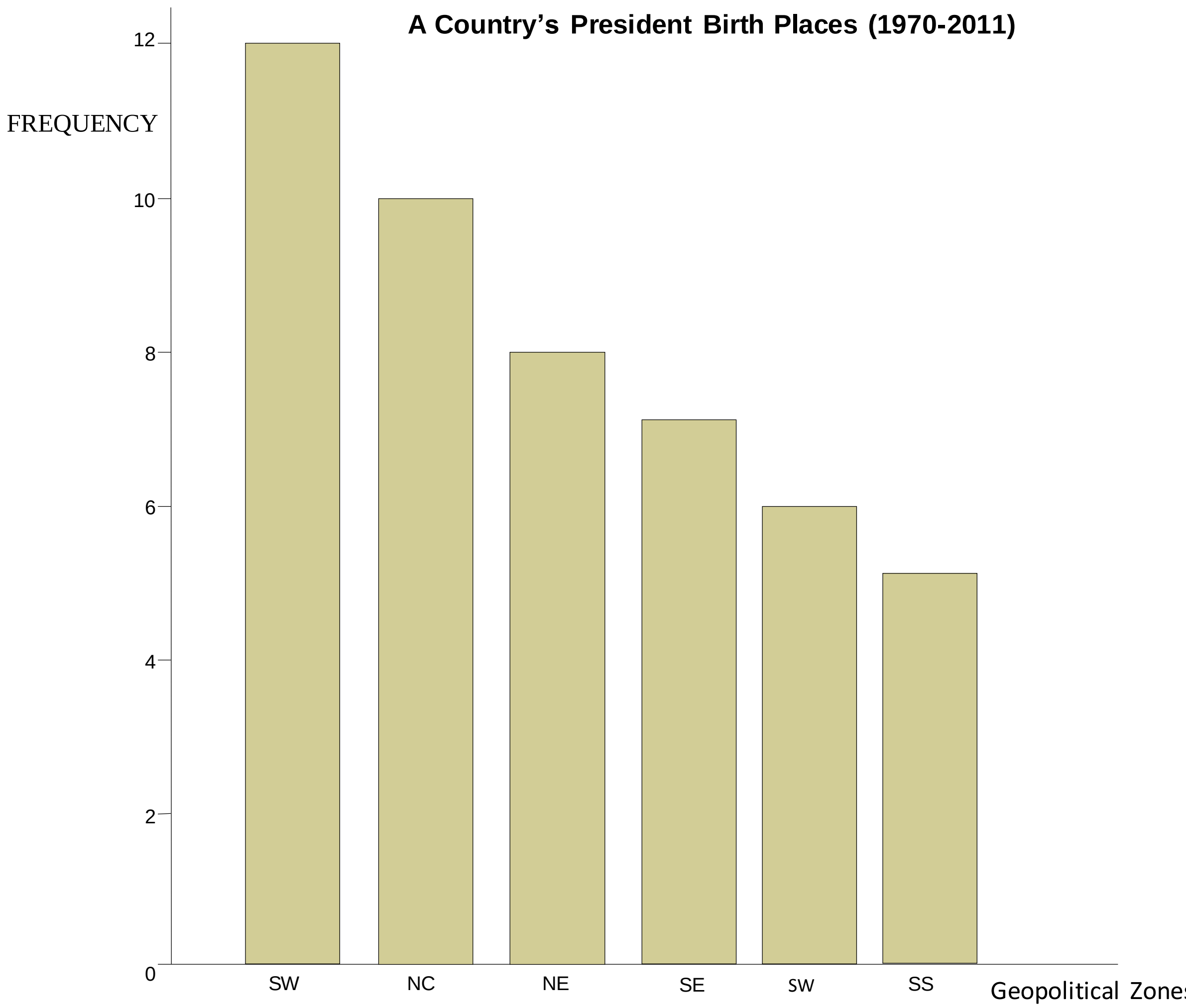
Example

The table below shows the number of people that have become the president of a country from its six geopolitical zones between 1940 and 2011.

Geopolitical Zone	Frequency
North Central (NC)	10
North East (NE)	8
North West (NW)	6
South East (SE)	7
South West (SW)	12
South South (SS)	5

Required: Present the data using the Pareto chart.

Solution



3.2.8 Pie-Charts

A pie chart is a circle divided by radial lines into sections (like slices of a cake or pie; hence the name) so that the area of each section is proportional to the size of the value represented. Each part of the sections is called a sector of a circle.

A pie chart can also be described as a disk divided into wedge-shaped pieces that are proportional to the frequencies or relative frequencies. Pie charts are sometimes called circular diagram. Each sector is quantified with the use of degree of the circle is proportional to the quality the sector represents; and the sum of all the angles (for all sector) gives 360^0. This could serve as a measure of accuracy in preparing pie charts.

A pie chart is particularly useful where it is desired to show the relative proportion of the values or variables that make up a single overall total.

Example 1

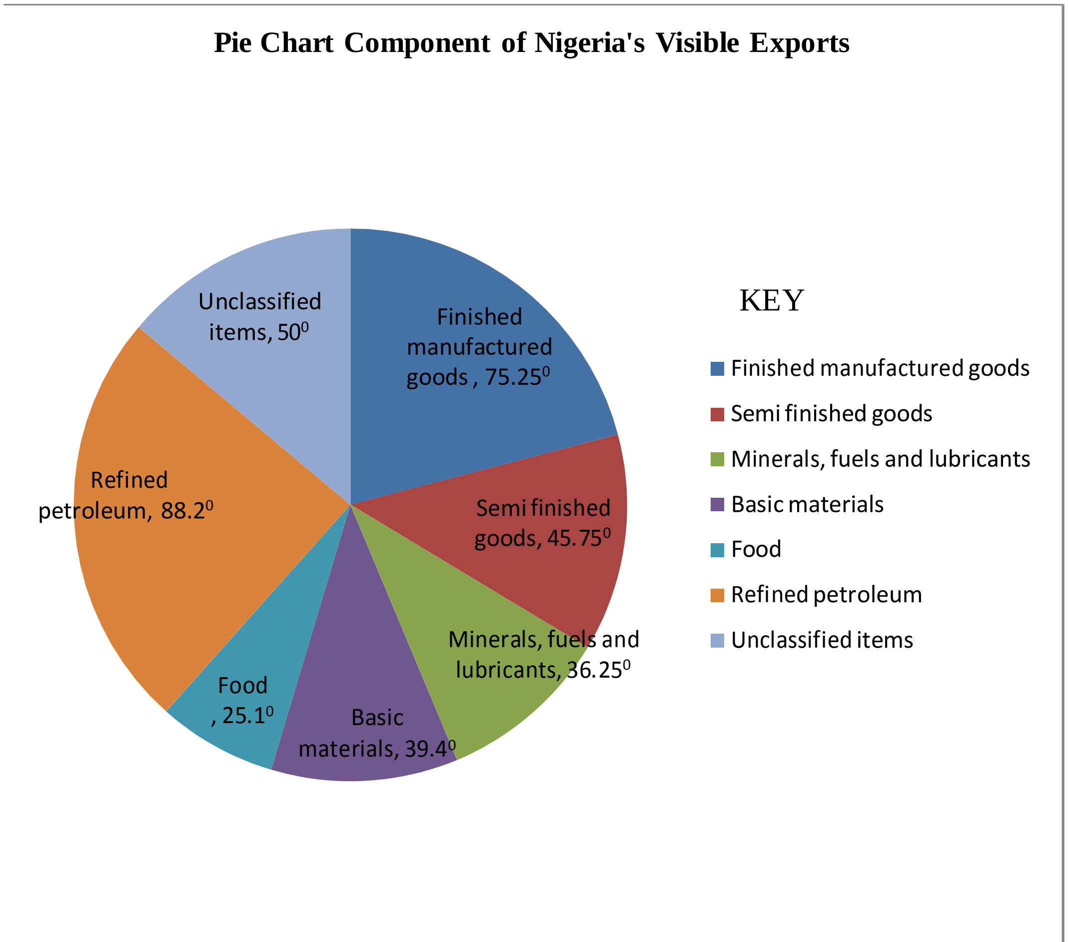
Given the components of the Nigeria's Visible Exports is a particular year in the table below: -

Components	Value (₦' Million)
Finished manufactured goods	150.5
Semi finished goods	91.5
Minerals, fuels and lubricants	72.5
Basic materials	78.8
Food	50.2
Refined petroleum	176.5
Unclassified items	100.0

Required: Present the information above using pie chart.

Solution

Components	Value (₦' Million)	Angle of each sector
Finished manufactured goods	150.5	$\frac{150.5}{720} \times \frac{360}{1} = 75.25^0$
Semi finished goods	91.5	$\frac{91.5}{720} \times \frac{360}{1} = 45.75^0$
Minerals, fuels and lubricants	72.5	$\frac{72.5}{720} \times \frac{360}{1} = 36.25^0$
Basic materials	78.8	$\frac{78.8}{720} \times \frac{360}{1} = 39.40^0$
Food	50.2	$\frac{50.2}{720} \times \frac{360}{1} = 25.10^0$
Refined petroleum	176.5	$\frac{176.5}{720} \times \frac{360}{1} = 88.25^0$
Unclassified items	100.0	$\frac{100}{720} \times \frac{360}{1} = 50.0^0$
Total	720	360.0^0



Example 2

The allocation of money to the manager sectors of the economy is given as follows:

Security	3x%
Education	25%
Health	15%
Sports	5.5%
Agriculture	12.5%
Transport	10%
Others	x%

Required

(a) Find the value of x in %?
(b) Draw a pie chart to illustrate the information.
(c) If N500 million is spent on education, how much is spent on:
    (i) Security
    (ii) Sports and transport
    (iii) The entire sector

Solution

(a) $3x + 25 + 15 + 5.5 + 12.5 + 10 + x = 100$

$68 + 4x = 100$

$4x = 100 - 68$

$4x = 32$

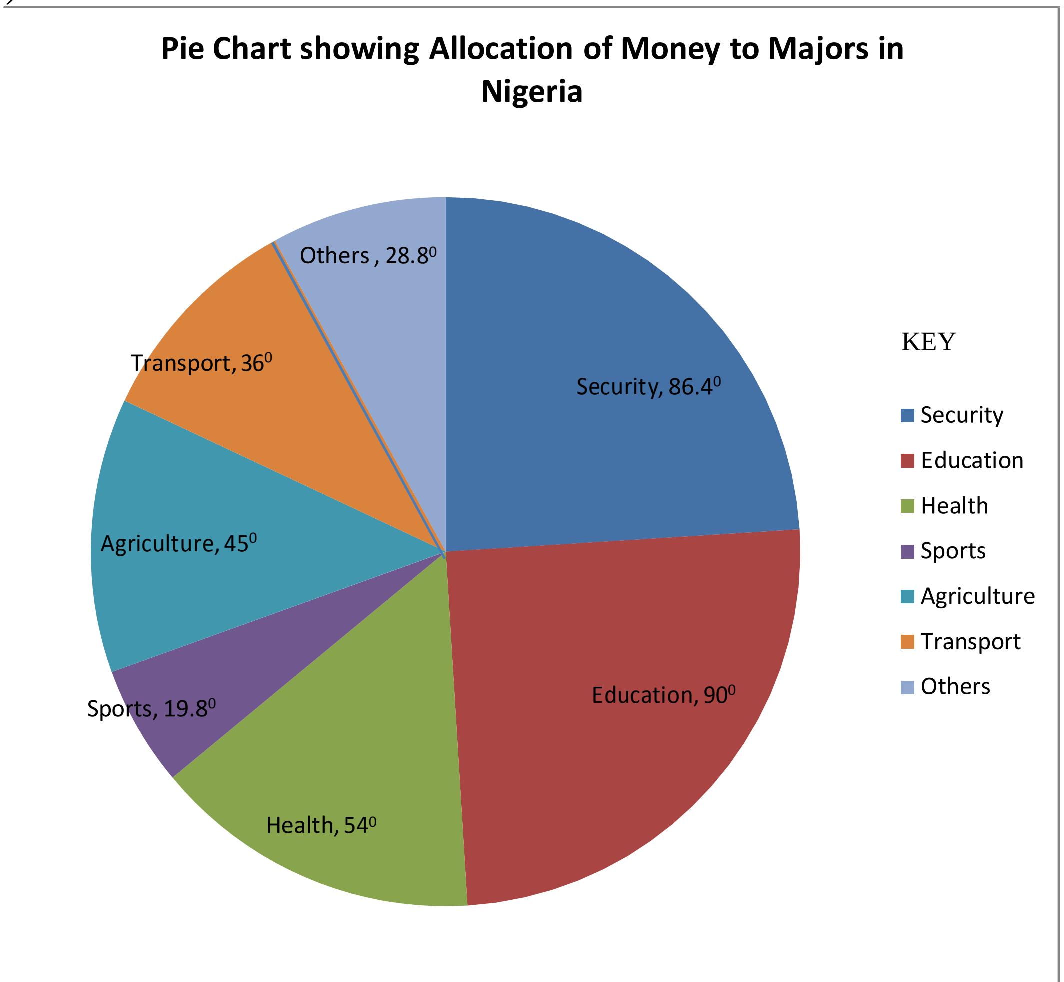
$x = \frac{32}{4}$

$x = 8\%$

ThRfRfofR, $3x = 3 \times 8 = 24\%$

Sectors	**%**	**Angle of each sector**
Security	24	$\frac{24}{100} \times 360 = 86.4^0$
Education	25	$\frac{25}{100} \times 360 = 90^0$
Health	15	$\frac{15}{100} \times 360 = 54^0$
Sports	5.5	$\frac{5.5}{100} \times 360 = 19.8^0$
Agriculture	12.5	$\frac{12.5}{100} \times 360 = 45^0$
Transport	10	$\frac{10}{100} \times 360 = 36^0$
Others	8	$\frac{8}{100} \times 360 = 28.8^0$
		360.0^0

(b)



(c) If N500 million is spent on education,

25% = N500 million

$1\% = \frac{N500}{25}$ million

= 20 million

(i) Security = 24% = N200 million × 24 = N480 million

(ii) Sports and transport = 5.5 + 10% = 15.5%

= N200 million × 15.5 = N310 million

(iii) The entire sector = 100%

= N20 million × 100 = N200,000 million

3.2.9 Z-Charts

A Z-chart is simply a graph that extends over a single year and incorporates:

(a) Individual monthly figures;
(b) Cumulative figure for the period; and
(c) The moving annual total.

It takes its name from the fact that the three curves (obtainable from a, b, c above) tend to look he letter ‘Z’.

Example

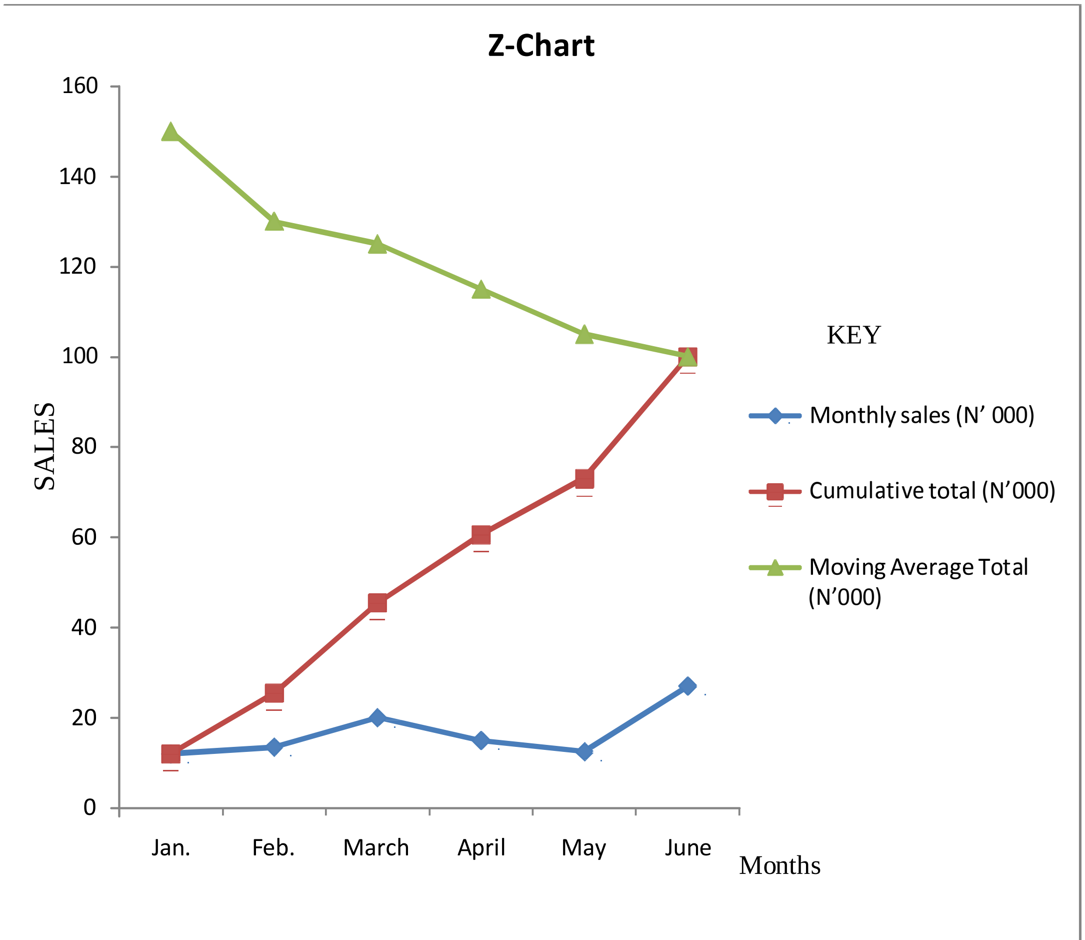
Given the sales value and the moving area total of a supermarket in N’ 000 as follows:

	Jan.	**Feb**	**March**	**April**	**May**	**June**
Sales volume	12	13.5	20	15	12.5	22.27
Moving Average	150	130	125	115	105	100

Required:

(a) Prepare the table for the cumulative total
(b) Construct the Z-chart.

Months	**Monthly sales (₦’ 000)**	**Cumulative total (₦’000)**	**Moving Average Total (₦’000)**
Jan.	12	12	150
Feb.	13.5	25.5	130
March	20	45.5	125
April	15	60.5	115
May	12.5	73.0	105
June	27	100	100



SELF ASSESSMENT EXERCISE 2

1. Given the votes of ANPP in six geo-political zones n Nigeria as: -

Zones	South East	South West	North Central	North East	North West	South South
% of votes	20%	25%	x%	19%	6%	2x%

(a) Find x in degree and percentage?
(b) Find the number of votes for ANPP in each region if 4.8 million votes were casted in the North-West region?
(c) Present the data in the table in a pie-chart.

4.0 CONCLUSION

The unit examined the meaning of charts and bar charts. It also highlighted the various forms of bar charts used in statistical analysis. Other forms of charts apart from bar charts (pie charts and Z-charts) were equally examined.

5.0 SUMMARY

Bar charts are descriptive representation of data with the use of bars (vertical or horizontals). There are various types of bar charts in statistics. The commonest ones include vertical bar charts, horizontal bar charts, component bar charts, percentage component bar charts, multiple bar charts etc. Apart from bar charts, other forms of charts are pie chart and Z-charts. While pie chart present data in a circular space, Z-chart present data with the use of three forms of data (current figures cumulative figures and total moving averages) to present three different curves that jointly bring the ‘Z’ shape.

6.0 TUTOR MARKED ASSIGNMENT

1. The expenditure of a local government council in 1984 is given as:

Education	₦ 212,000
Health	₦ 94,000
General	₦ 102,000
Development	₦ 72,000
Roads	₦ 81,000
Social welfare	₦ 30,000
Others	₦75,000

Required: Present the data in a pie chart.

2. The table below shows the volume of a country’s food crop export in tonnes between 1990 and 1993:

Years	**Production in Tonnes**			
	1990	1991	1992	1993
Groundnut	600	600	700	900
Palm oil	500	510	520	600
Soyabeans	710	720	800	1000

Required: Present the following

(a) Component bar cart
(b) Multiple bar chart

(c) Percentage bar chart

3. The grade of 170 students in an examination is given as thus: -

Grades	A	B	C	D	E	F
Frequency	15	25	20	45	30	35

Required: Present the information using:

(a) Simple horizontal bar chart
(b) Simple vertical bar chart
(c) Pareto chart

4. The following shows the number of stores that purchase swam water in Jos, Akure, Abuja and Akwa-Ibom respectively.

Jos	25,000
Akure	10,000
Abuja	20,000
Akwa-Ibom	5,000

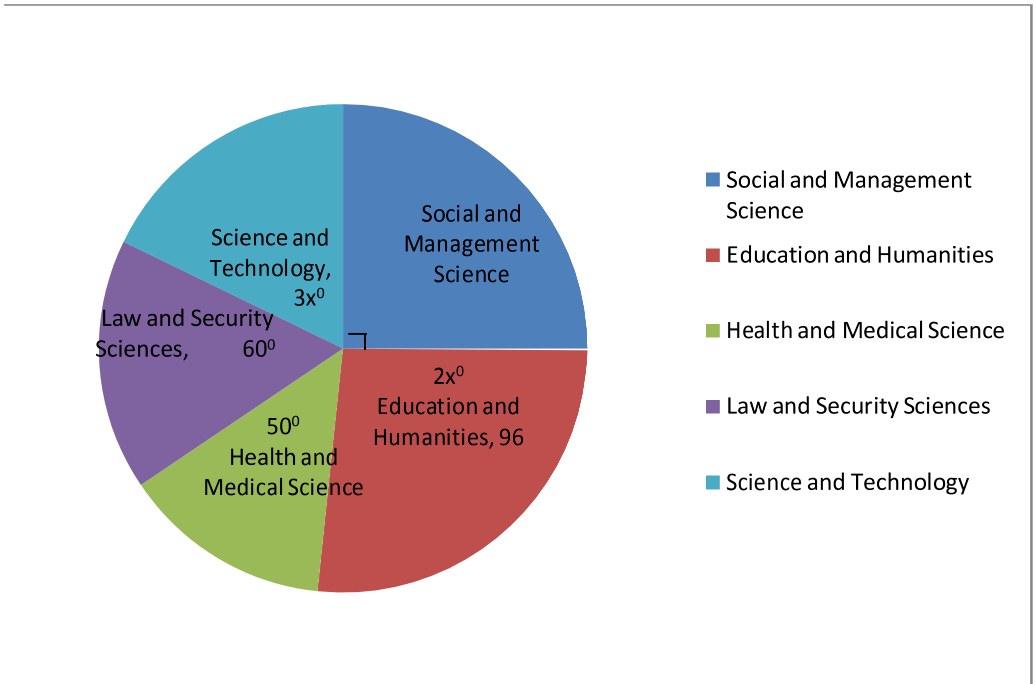
Present the above purchases of swam water information on a pie chart.

5. Agada food is projecting the demand for its product as follows:

Year	**Demand**
1981	800
1982	1200
1983	1100
1984	1400
1985	1600
1986	1700

Present the data in a simple horizontal bar chart.

6. Given the distribution of students in various colleges/faculty I a university by pie chart drawn;



(a) Find x in degree, hence, find the proportion of students in:
   (i) Science and technology
   (ii) Education and humanities

(b) If there are 360 students in the school of education and humanities, determine the number of students in;
   (i) Each of the other colleges/schools.
   (ii) The entire university.
   (iii) In the school that has the least population and the school that has the highest population.

7.0 REFERENCES/FURTHER READINGS

Frank, O. and Jones, R. (1993). Statistics. Pitman Publishing, London.

Levin, R.I. (1988).Statistics for Managers, Eastern Economy Edition, Prentice-Hall of India Private Limited.

Loto, M.A., Ademola, A.A. and Toluwase, J.A. (2008). Statistics Made Easy. Concept Publication.

Spiegel, M. and Stephen, L. T. (2004).Statistics. Schaums Outline Series, Tata McGraw-Hill Limited, New Delhi.

Neil, A. Weiss (2008). Introductory Statistics. Pearson International Edition (8^{th} Edition). Addison Wesley Publisher, United State.

UNIT 4

HISTOGRAMS AND CURVES

Table of Contents

1.0 INTRODUCTION

Apart from tables and charts, curves are also used in representation of data. A special type of bar charts – Histogram is also a very good way of representing data. A histogram is constructed by drawing rectangles for each class or variable. The height of each rectangle is the frequency for relative frequency of the class. The width of each rectangle should be the same and the rectangles should touch each other. Histogram may be used for both grouped and ungrouped data. A number of curves are also used in statistics to represent data. Some of them include cumulative frequency curves, Lorenz curves etc.

2.0 OBJECTIVES

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

(i) Construct histogram and frequency polygon of different forms of data;
(ii) Construct cumulative frequency curves for given set of data;
(iii) Use Lorenz curve and pictogram to represent data.

3.0 MAIN CONTENT

3.1 Histogram and Frequency Polygon

Histogram and frequency polygon are usually drawn together. While histograms are vertical bar charts joined together, frequency polygon is the curve obtained by joining the midpoint of each bars of a histogram. A histogram, otherwise known as frequency histogram consists of a set of rectangles having bases on a horizontal axis (the X-axis) with the centers at the class marks and lengths equal to the class interval sizes and area proportional to the class frequencies. A frequency polygon is a line graph of the class frequency plotted against the class mark. It can be obtained by connecting the midpoints of the top of the rectangle in the histogram.

3.1.1 Histogram and Frequency Polygon of Ungrouped Data

For ungrouped data, the X-axis represents the variable or the item while the Y-axis represents the frequency. The midpoint of each bars are joined together to form the frequency polygon.

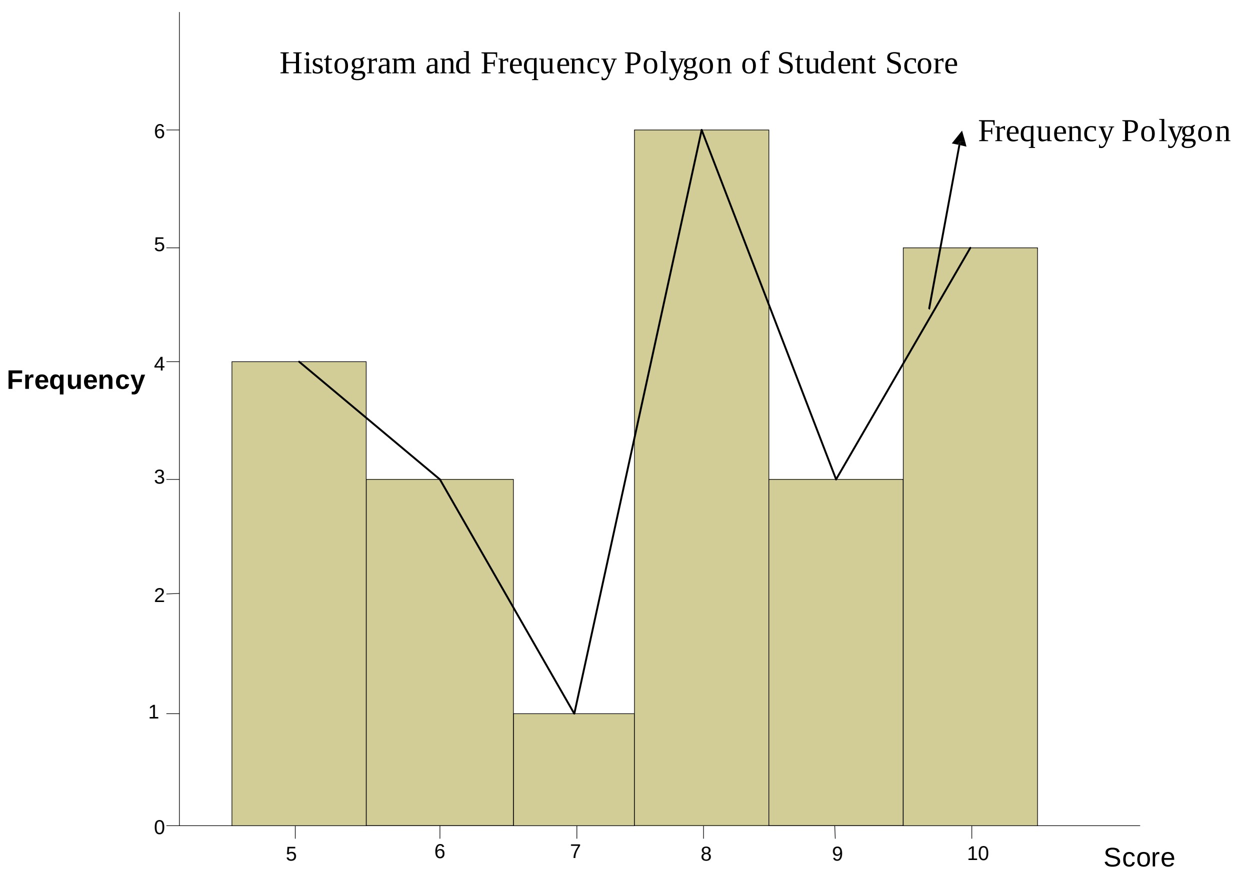
Example

Given the scores of students in a test as follows:

Scores (x)	5	6	7	8	9	10
Frequency (f)	4	3	1	6	3	5

Represent the data in histogram and frequency polygon.

Solution



3.1.2 Histogram and Frequency Polygon of Grouped Data

For grouped data, the vertical axis represents the frequency while class boundaries are stated on the horizontal axis.

Example

Given the scores of student in an examination as:

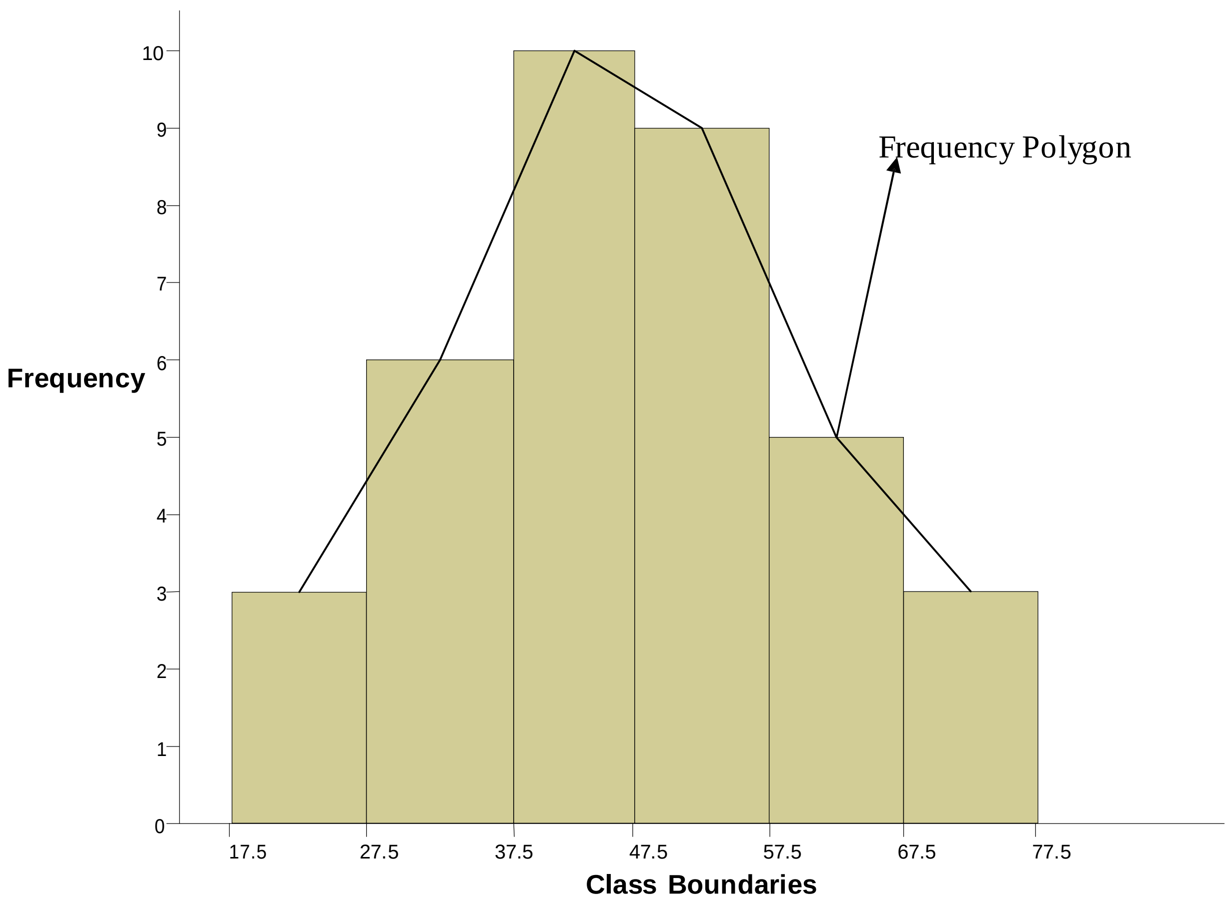
Marks (x)	18–27	28–37	38–47	48–57	58–67	68–77
Frequency (f)	3	6	10	9	5	3

(a) Represent the information using histogram and frequency polygon.

(b) Construct the histogram and frequency polygon using the frequency density.

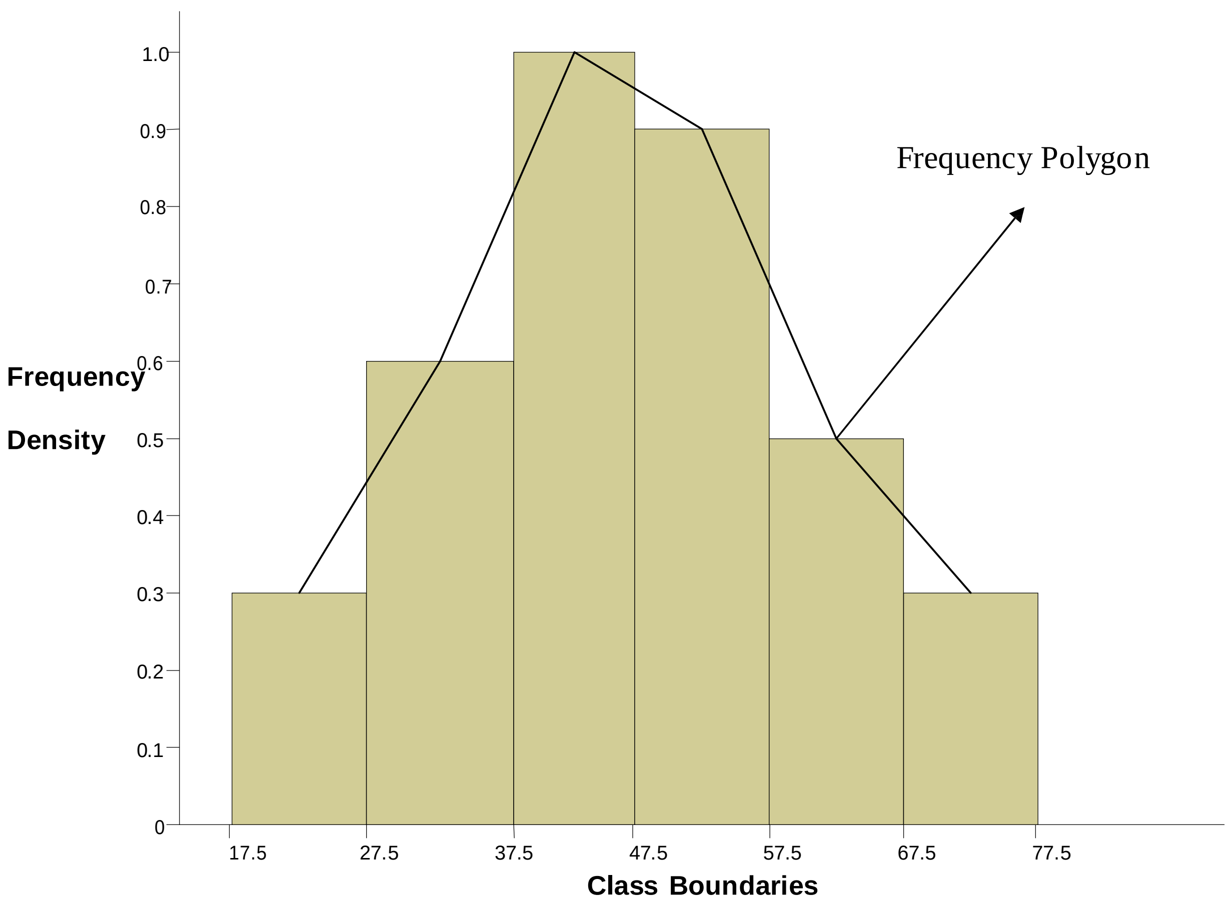
Solution

a. Histogram and Frequency Polygon of Grouped Data



Marks	Frequency	Frequency Density	Class Boundaries
18 – 27	3	$\frac{3}{0} = 0.3$	17.5 – 27.5
28 – 37	6	$\frac{6}{0} = 0.6$	27.5 – 37.5
38 – 47	10	$\frac{0}{0} = 1.0$	37.5 – 47.5
48 – 57	9	$\frac{9}{0} = 0.9$	47.5 – 57.5
58 – 67	5	$\frac{5}{0} = 0.5$	57.5 – 67.5
68 – 77	3	$\frac{3}{0} = 0.3$	67.5 – 77.5

Note: Frequency Density = $\frac{\text{Frequency}}{\text{C a Interva}}$



3.1.3 Relative Frequency Histogram

The relative frequency histogram is the graph that displays the classes on the horizontal axis and the relative frequencies of the classes of the vertical axis. The relative frequency of each class is represented by a vertical bar whose height is equal to the relative frequency of the class.

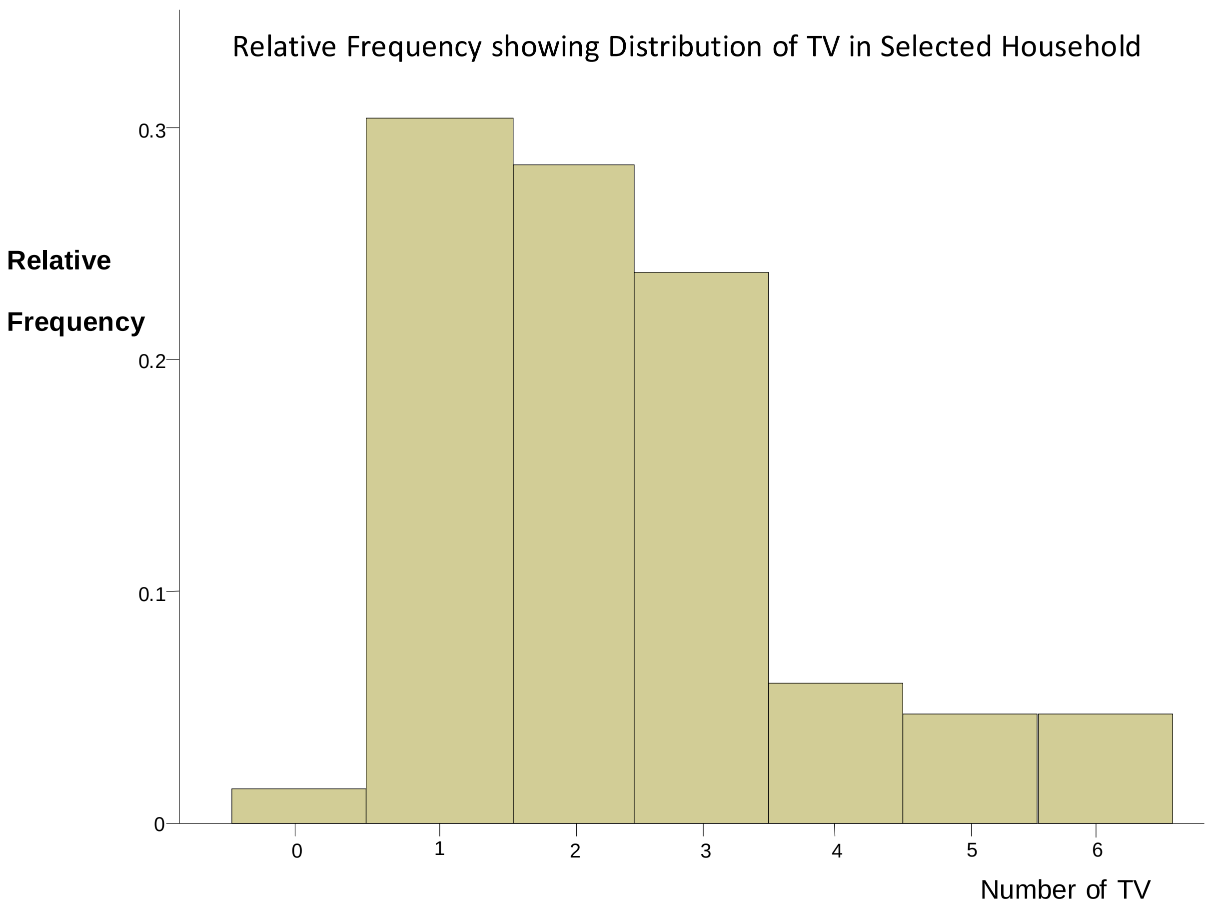
Example

Given the frequency of number of televisions in a given household as follows: -

No of TV	0	1	2	3	4	5	6
Frequency (f)	1	16	14	12	3	2	2

Required: - Construct the relative frequency distribution for number of TV sets?

Solution



Note: Relative frequency = $\frac{\text{Frequency}}{\text{Tota Frequency}} \times \frac{00}{\ }$

∴ The relative frequency table is given as:

No of TV	Frequency	Relative Frequency
0	1	0.2
1	16	0.32
2	14	0.28
3	12	0.24
4	3	0.06
5	2	0.04
6	2	0.04
	t = so	**1.00**

3.1.4 Percentage Relative Frequency

This is the graph that displays the classes on the horizontal axis and the percentage relative frequency of the classes on the vertical axis. When the midpoints of the bars are joined together we obtain the percentage relative frequency polygon.

Example

Given the distribution of students' scores in an examination as follows:

Scores	1 – 20	21 – 40	41 – 60	61 – 80	81 – 100
Frequency	1	16	14	12	7

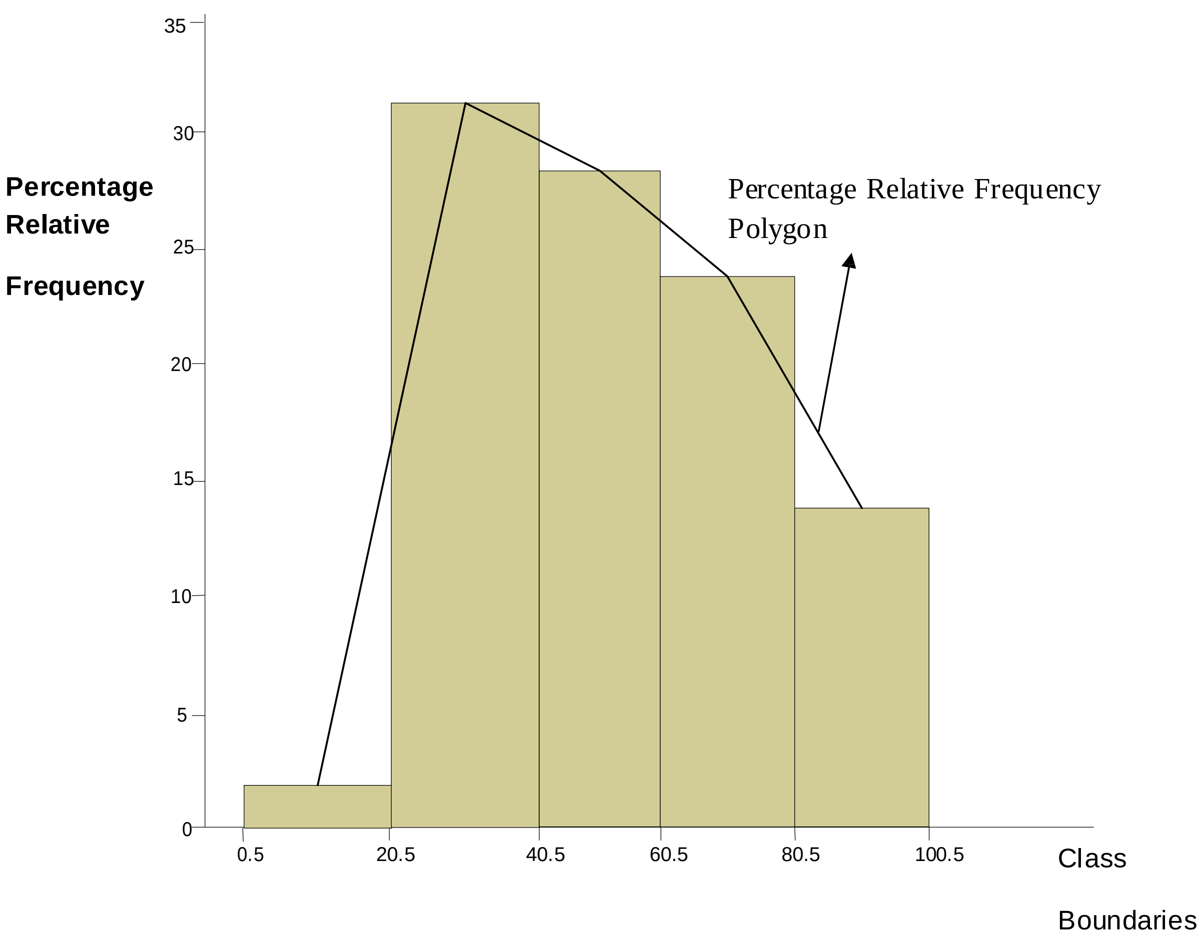
Required: Present the data using percentage relative frequency polygon.

Solution

Percentage relative frequency = RelatiR ffReqR C $\times \frac{00}{\ }$ of

$$\frac{\text{Frequency}}{\text{Tota Frequency}} \times \frac{00}{\ }$$

Classes	Frequency	Class Boundaries	Relative Frequency	Percentage Relative Frequency
1 – 20	1	0.5 – 20.5	0.02	2.0
21 – 40	16	20.5 – 40.5	0.32	32.0
41 – 60	14	40.5 – 60.5	0.28	28.0
61 – 80	12	60.5 – 80.5	0.24	24.0
81 – 100	7	80.5 – 100.5	0.14	14.0
	t = so			



CLASS ASSESSMENT EXERCISE 1

1. Given the age distribution of pensioners as:

Ages	50 – 54	55 – 59	60 – 64	65 – 69	70 – 74	75 – 79
Frequency	12	10	8	14	16	10

Present the data using:

(a) Histogram and frequency polygon
(b) Relative frequency polygon
(c) Relative frequency histogram
(d) Percentage relative frequency polygon

2. The scores of 100 students in a test is given as: -

Scores	0	1	2	3	4	5	6	7	8	9	10
Frequency	2	10	7	8	5	10	15	10	7	5	21

Present the data using:

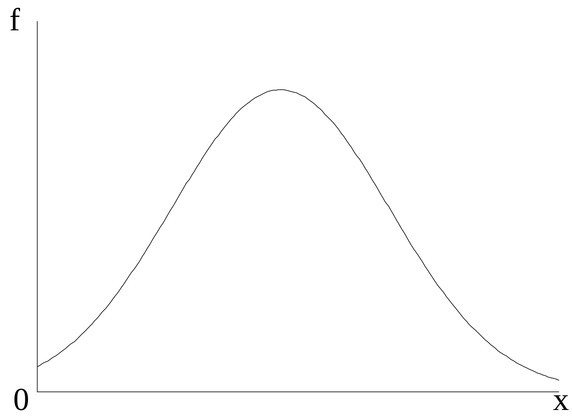
(a) Histogram
(b) Frequency polygon
(c) Relative frequency histogram

3.2 Frequency Curves

Collected data can usually be considered as belonging to sample drawn from a large population. Since so many observations are available in the population, it is theoretically possible to choose intervals very small and still have sizeable numbers of observations falling within each class. Thus, one would expect the frequency polygon or relative frequency polygon for a large population to have many small, broken line segments that they closely approximate curves, which we call frequency curves or relative frequency curves respectively. It is reasonable to expect that such theoretical curves can be approximated by smootheing the frequency polygons or relative frequency polygons of the sample, the approximation improving as the sample size is increased. For this reason, a frequency curve is sometimes called a frequency polygon.

Frequency curves arising in practice take on certain characteristics shapes, as shown below:

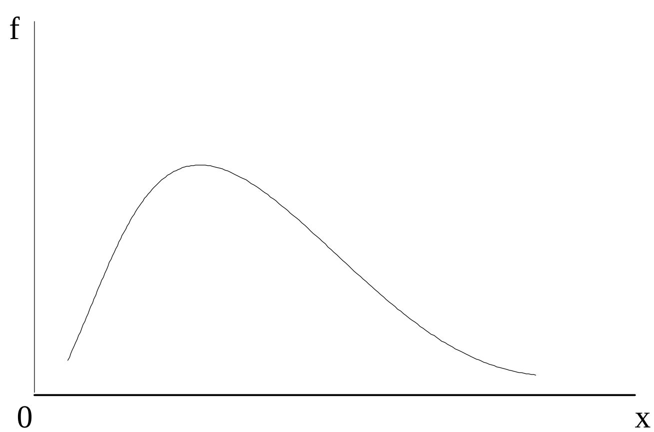
(a)



Symmetrical or Bell-Shaped

This is characterized by the fact that observations equidistant from the central maximum have the same frequency. An important example is the normal curve

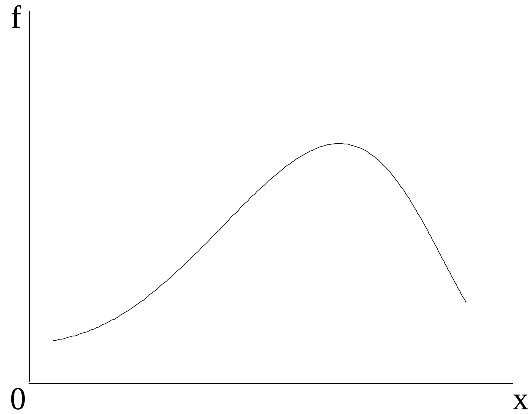
(b)



Skewed to Result (Positive Skewness)

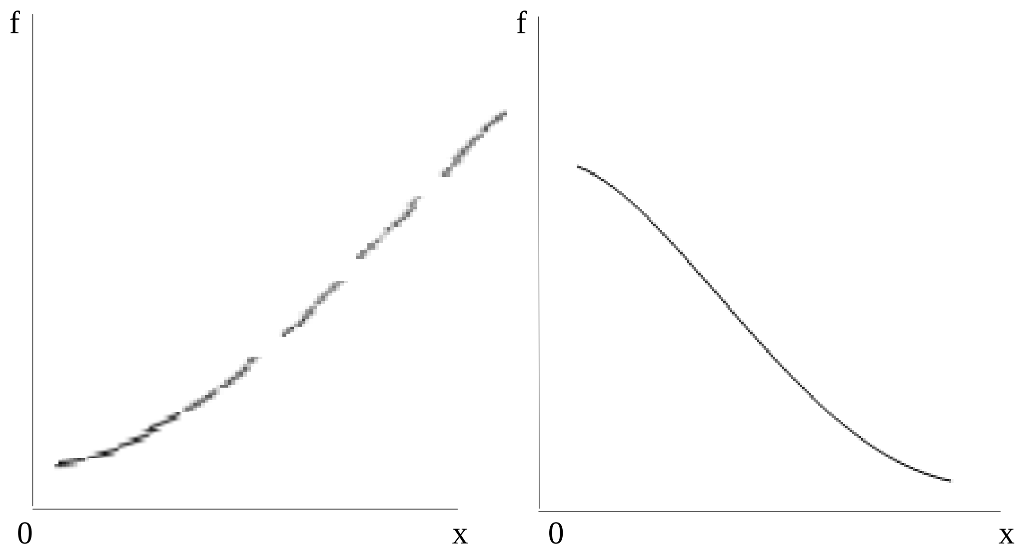
In the moderately asymmetrical or skewed, frequency curves, the tail to one side of the central maximum is longer than the other. If the longer tail occurs to the right, the curve is said to be skewed to the right or to have positive skewness, while if the reverse is true, he curve is said to be skewed to the left or have a negative skewness.

(c)

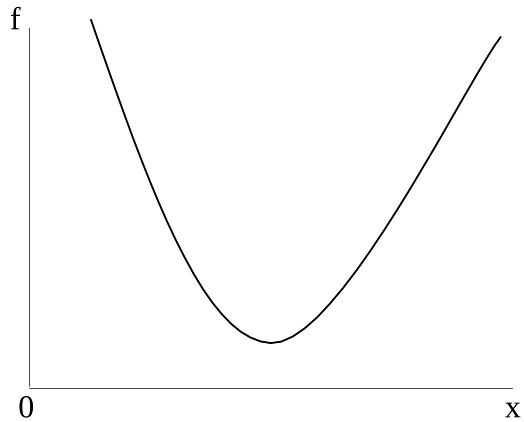


Skewed to the Left or Negative Skewness

(d) J shaped or Reverse J-Shaped: - In this type of curves, the maximum occur at one end or the other.

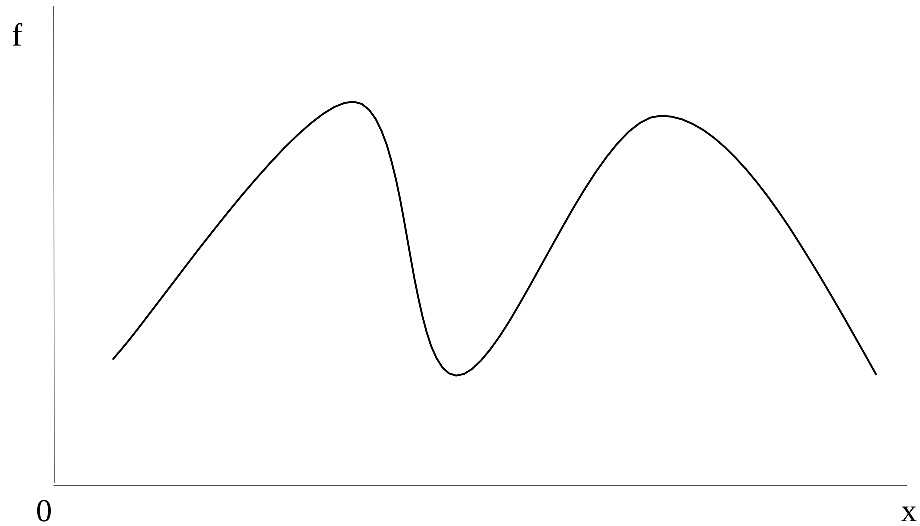


(e) U-shaped

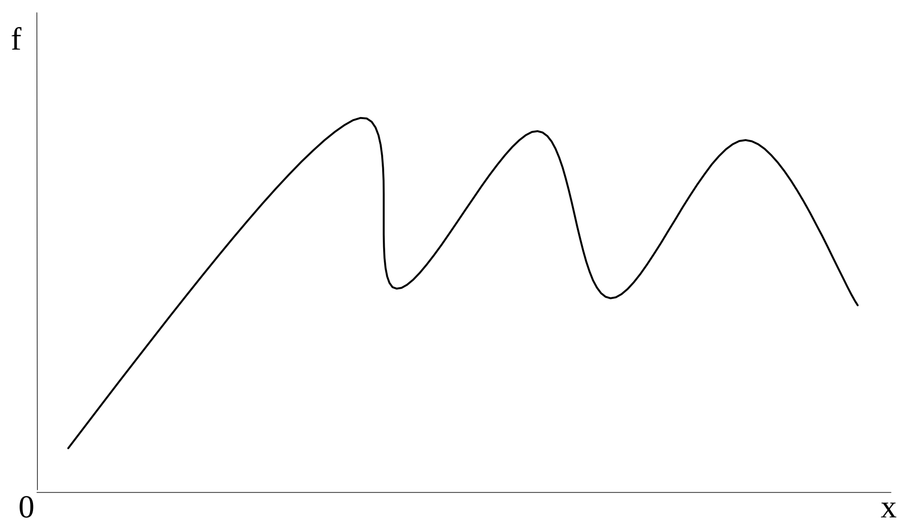


The frequency curve has maxima at both ends.

(f) Bimodal frequency Curve has two maxima or modal points.



(g) Bimodal frequency curve has more than two maxima or modal points



SELF ASSESSMENT EXERCISE 2

1. What are frequency curves?
2. Distinguish clearly between frequency curves and frequency polygons.
3. With the aid of graph, explain each of the following concepts.
   (a) Symmetrical or Bell-shaped curve
   (b) U-shaped curve
   (c) Reverse J-shape curve
   (d) Multi-modal curves

3.3 Cumulative Frequency

The total frequency of all values less than the upper class boundary of a given class interval is called cumulative frequency up to and including that class interval. A table presenting such cumulative frequencies is called cumulative frequency distribution, cumulative frequency table or briefly a cumulative distribution. A graph showing the cumulative frequency less than any upper class boundary plotted against the upper class boundary is called a cumulative frequency curve or ogive curve. In drawing the ogive curve, the upper class boundary is placed at the horizontal axis and the cumulative frequency at the vertical axis.

Example

Given the scores of 100 students in an examination as follows: -

Scores	1 – 10	11 – 20	21 – 30	31 – 40	41 – 50	51 – 60	61 – 80
Frequency	10	15	25	5	20	35	40

Required: -

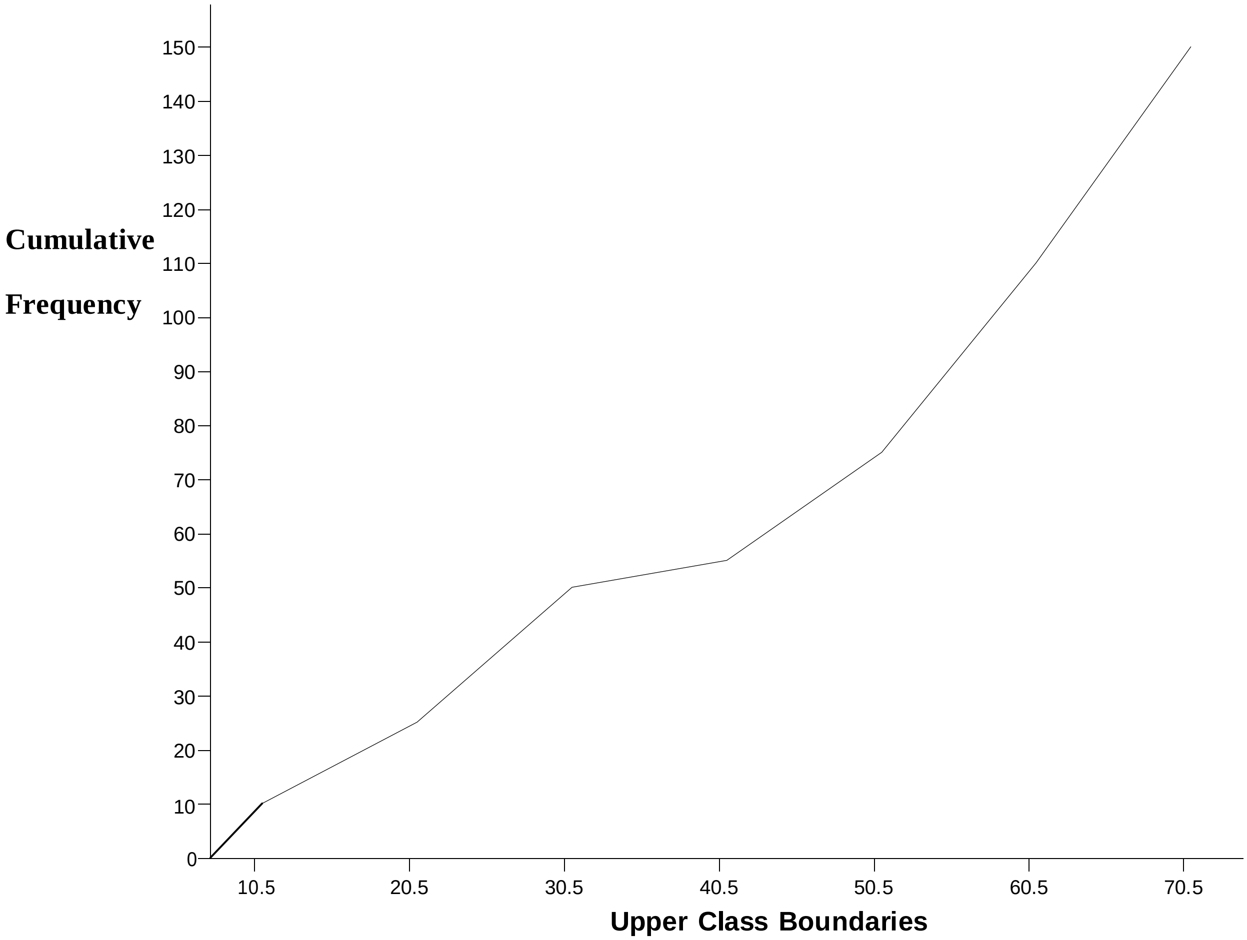
(a) Construct the class boundary cumulative frequency, the percentage cumulative and the relative cumulative frequency tables for the data.
(b) Construct the cumulative frequency curve or ogive curve for the data.
(c) Construct the percentage cumulative frequency curve for the data.
(d) Construct the relative cumulative frequency curve for the data
(e) What do you notice about the curves in (b), (c) and (d) above.

Solution

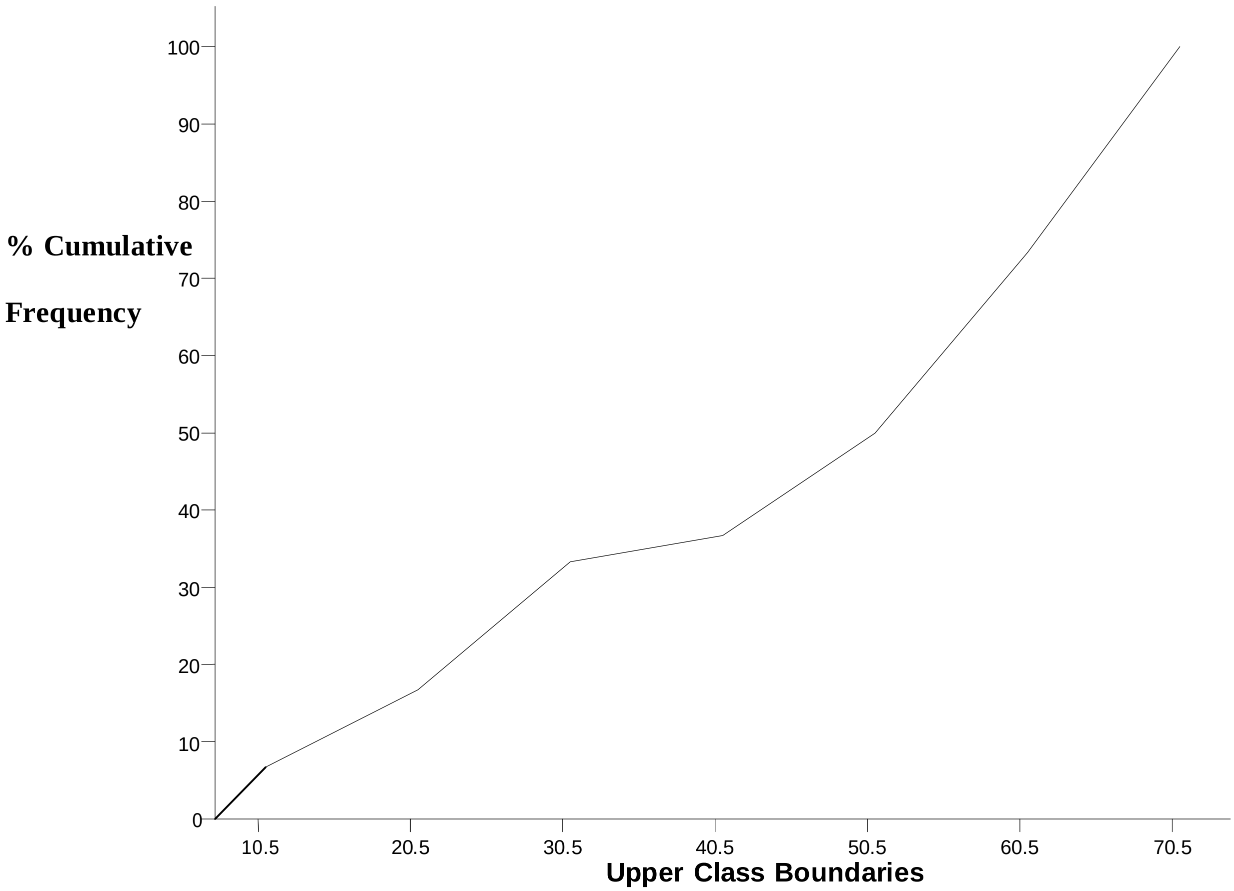
a.

Scores	**Class boundaries**	**Freq.**	**Cum. Freq.**	**% Cum. Freq.**	**Rel. Freq**	**Rel. Cum. Freq.**
1 – 10	0.5 – 10.5	10	10	6.7	0.07	0.07
11 – 20	10.5 – 20.5	15	25	16.7	0.10	0.17
21 – 30	20.5 – 30.5	25	50	33.3	0.17	0.34
31 – 40	30.5 – 40.5	5	55	36.7	0.03	0.37
41 – 50	40.5 .50.5	20	75	50.0	0.13	0.50
51 – 60	50.5 – 60.5	35	110	73.3	0.23	0.73
61 – 70	60.5 – 70.5	40	150	100	0.27	1.00

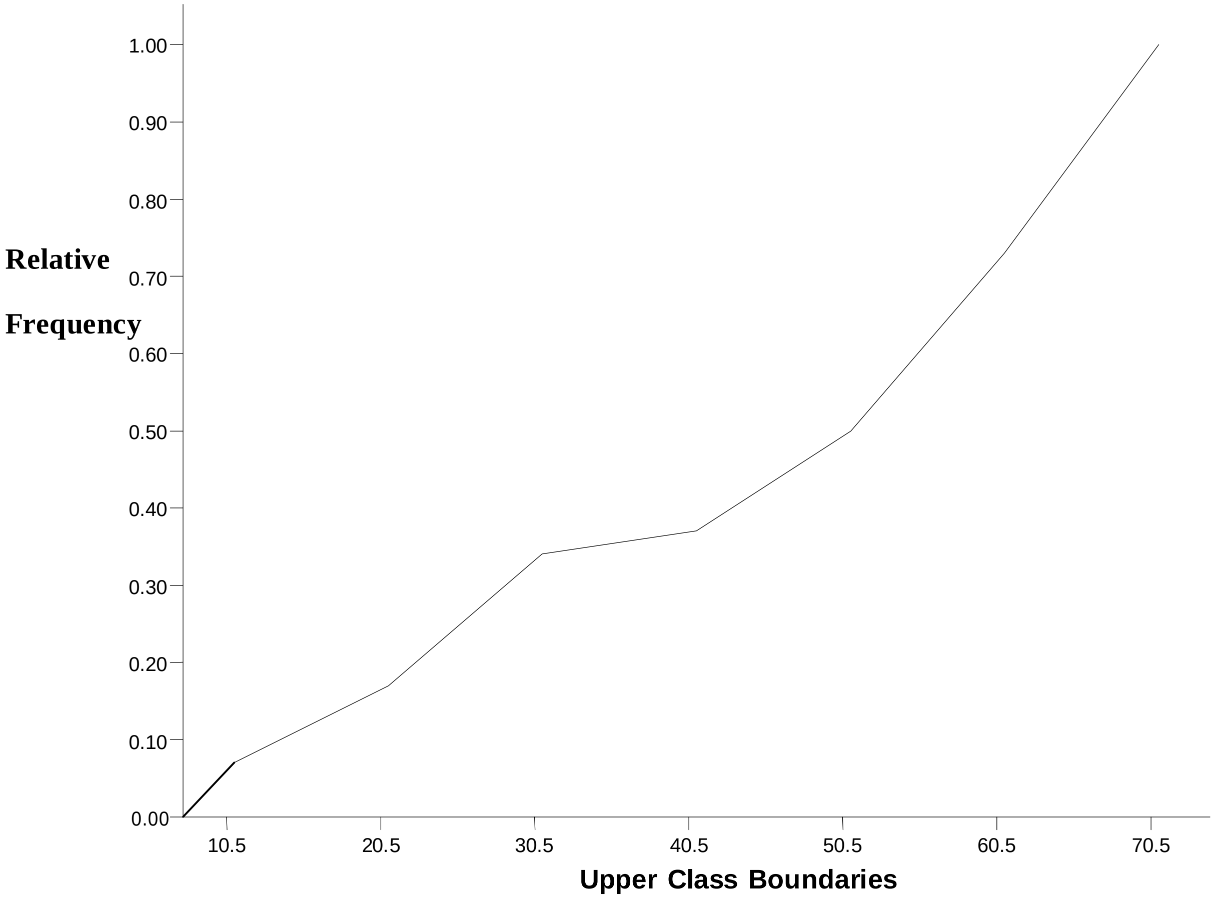
b. Cumulative Frequency Curve



c. Percentage Cumulative Frequency Curve



d. Relative Cumulative Frequency Curve



e. All the curves have the same shape. This is because the construction is based o the same principle (cumulative). The only difference is the denomination use for constructing the cumulative. In the first graph, absolute cumulative frequency is used; percentage cumulative is used the second while relative cumulative frequency is used in the third.

CLASS ASSESMENT EXERCISE 3

1. Given the age distribution of workers in an organization as follows:

Age Group	21 – 30	31 – 40	41 – 50	51 – 60	61 – 70	71 – 80
Freq.	5	30	23	25	15	2

a. Represent the data by using:
   (i) Cumulative frequency curve
   (ii) Percentage cumulative frequency curve
   (iii) Relative cumulative frequency curve
b. Compare and construct the curves in a (i), (ii) and (ii).

3.4 Lorenz Curves and Pictograms

Lorenz-curve is a graphical method for demonstrating the disparity between two economic phenomena. It gives visual impression of the degree of inequality or uneven distribution in the economy.

Like other types of graph, a Lorenz curve is basically used to facilitate comparison of data in income, production, sales etc; with a view of showing the extent of concentration of each phenomenon in the hands of either the individuals, or firms or industries. It can also be used to show the degree of inequality in the distribution of income or wealth of the nation. It can also be used to show the extent to which profit of various companies vary. Lorenz curves can also be used to show the extent to which taxes range with income of population and variations in the output of firms of different sizes.

In constructing a Lorenz curve, the following has to be done: -

(i) Compute the cumulative frequency of the variables;
(ii) Compute the cumulative percentage for the variable;
(iii) Plot the cumulative percentage of one variable against the other; and
(iv) Draw the line of equal distribution to bring out very clearly the extent of divergence from a uniform distribution.

Example

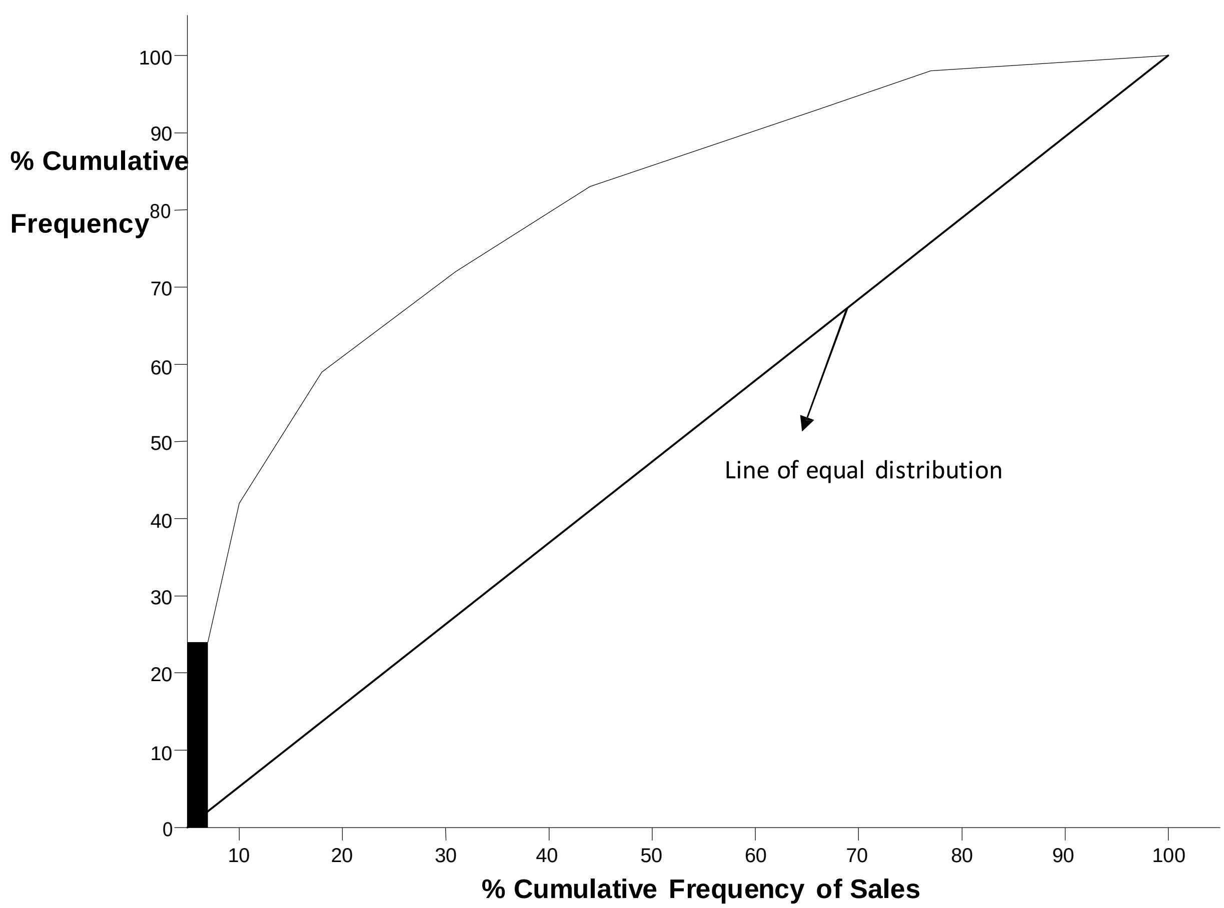
Given the table below: -

Investment (₦)	No of Bookshops	Sales Value (₦' 000)
0 – 999	96	120
1000 – 1999	72	250
2000 – 2999	67	305
3000 – 3999	52	460
4000 – 4999	46	495
5000 – 5999	39	580
6000 – 6999	21	626
7000 – 7999	8	852

Required: - Present the Lorenz curve of Bookshop by size of investment and sales value?

Solution

Investment (₦)	No of Bookshops	Sales (₦' 000)	Cumulative of Bookshop		Cumulative Sales	
			Number	%	₦' 000	%
0 – 999	96	120	96	24	250	7
1000 – 1999	72	250	168	42	370	10
2000 – 2999	67	305	235	59	675	18
3000 – 3999	52	460	387	72	1135	31
4000 – 4999	46	495	333	83	1630	44
5000 – 5999	39	580	372	93	2210	60
6000 – 6999	21	626	393	98	2836	77
7000 – 7999	8	852	401	100	3688	100



Pictogram: - This is the use of shapes, figure or objects to represent information or data. It entails the use of drawing of objects to represent items in a data. The pictures or diagrams are used in such a way as to show the degree of occurrence of each variable in an observation over a period of time.

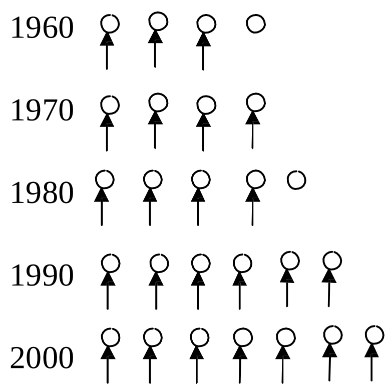
Example

The population of a country over the last 4 decades is given as: -

Year	1960	1970	1980	1990	2000
Population (Million)	70	80	90	120	150

Required: - Present the data using pictogram?

Solution



Note:  = 20 million People, ○ = 10 million people

CLASS ASSESSMENT EXERCISE 3

1. The number of people available to vote in a particular state in Nigeria in the last fifty years is given as: -

Year	1960	1970	1980	1990	2000
Eligible Voters (Thousand)	70	90	100	1300	190

2. Given the information below:

Size of Holding	Number of Holdings	Total Holdings
Under ½ hectare	31	10
½ to under 1 hectare	24	17
1to under 5 hectare	7	18
5 to under 15 hectare	3	30
15 hectare and above	2	42

Required: - Present the Lorenz curve of number of holding by the total area?

4.0 CONCLUSION

The unit examines the various forms of histogram and cumulative frequency used in representation of statistical data. It also extends to other forms of representation of data such as Z-charts and pictogram.

5.0 SUMMARY

Histograms are special form of bar charts in which information or data are represented with the aid of jointed bars. They are of different forms and types. Histograms are usually drawn along ways with frequency polygon (curves joining midpoints of histogram bars). Frequency curves can be of different shapes such as symmetrical 'U'-shaped, J-shaped etc. They are meant to describe the relationship which exists between observations and heir frequencies.

Cumulative frequency curves are drawn with the use of cumulative frequency distribution table. It is constructed by joining the co-ordinates of cumulative frequency drawn in a co-ordinate plane with upper class boundary in the horizontal axis and the cumulative frequency in the vertical axis. Different categories of cumulative frequency curves (relative cumulative curve, percentage cumulative frequency curves etc) can also be drawn by changing the variables in the vertical axis.

Lorenz charts are drawn to establish disparity which exists between two or more variable. They give visual impression about the degree of disparity (divergence or convergence) in the variables. To show pictogram distribution (especially for discrete variables), pictogram can be used.

6.0 TUTOR MARKED ASSIGNMENT

1) The table below shows a frequency distribution of salaries per annum in Naira of 100 employees at PPMD company: -

Salaries (₦'000)	Number of Employees
50 – 59.99	8
60 – 69.99	10
70 – 79.99	14
80 – 89.99	18
90 – 99.99	24
100 – 109.99	12
101 – 119.99	7
110 – 129.991	5
130 – 139.99	2

With reference to the table, construct:

(a) Cumulative frequency curve and frequency polygon; and
(b) Relative frequency curve.

2) The length of 40 leaves are recorded to the nearest millimeters in the table below:

Length (mm)	Frequency
118 – 126	3
127 – 135	5
136 – 144	9
145 – 153	12
154 – 162	5
163 – 171	4
172 – 180	2

Construct:

(a) Relative frequency distribution;
(b) Histogram;
(c) Relative histogram;
(d) Frequency polygon; and
(e) Relative frequency polygon.

3) The following table shows the diameters in centimeters of a sample of 40 balls manufactured by a company: -

17.38	17.29	17.43	17.36	17.35	17.31	17.26	17.37
17.28	17.37	17.36	17.24	17.42	17.36	17.39	17.35
17.45	17.36	17.42	17.28	17.25	17.33	17.34	17.52
17.33	17.30	17.32	17.39	17.36	17.41	17.36	17.44
17.32	17.37	17.31	17.35	17.29	17.34	17.30	17.40

(a) Construct a frequency distribution table of the diameter using the class interval: 17.01 – 17.9, 17.10 – 17.19, 17.20 – 17.29 …

(b) From the table, constructed above, construct:

(i) Histogram
(ii) Frequency polygon
(iii) Relative frequency distribution
(iv) Relative frequency polygon
(v) Relative frequency histogram
(vi) Cumulative frequency distribution
(vii) Percentage cumulative distribution
(ix) An ogive
(x) Percentage ogive.

4) Below is the value of sales of building materials to the different categories of consumers in Orisun state.

Year	**Values of Sales to:**		
	Non-Indigenous Contractors	Indigenous Contractors	Non Contractors
1978	325,000	300,000	105,000
1979	375,000	370,000	231,000
1980	292,000	212,000	197,000
1981	237,500	169,000	85,000
1982	465,000	175,000	112,000

Illustrate the figure by a means of:

(a) Multiple bar chart
(b) Component bar chart
(c) Percentage component bar chart

5) Given below is the monthly output of plastic cups by ESAKOT Nigeria Limited.

Month	Jan.	Feb.	Mar.	April	May	June	July	Aug.	Sept	Oct.
1991 (₦'000)	56	52	59	61	67	75	85	85	73	62
1992 (₦'000)	63	54	57	56	63	72	85	96	99	95

From the above information, construct a Z-chart for 1982. Comment on the chart.

7.0 REFERENCES/FURTHER READINGS

Frank, O. and Jones, R. (1993). Statistics. Pitman Publishing, London.

Levin, R.I. (1988).Statistics for Managers, Eastern Economy Edition, Prentice-Hall of India Private Limited.

Loto, M.A., Ademola, A.A. and Toluwase, J.A. (2008). Statistics Made Easy. Concept Publication.

Spiegel, M. and Stephen, L. T. (2004).Statistics. Schaums Outline Series, Tata McGraw-Hill Limited, New Delhi.

Neil, A. Weiss (2008). Introductory Statistics. Pearson International Edition (8^{th} Edition). Addison Wesley Publisher, United State.

Nwabuoku, P. O. (1986). Fundamental of Statistics; Roruna Book, Enugu, Nigeria.

Module 3: Basic Statistical Measures of Estimates

Unit 1: Measures of Central Tendency of Ungrouped Data

Unit 2: Measures of Central Tendency of Grouped Data

Unit 3: Measures of Dispersion

Unit 4: Measures of Partition

UNIT 1

MEASURES OF CENTRAL TENDENCY OF UNGROUPED DATA

Table of Contents

1.0 INTRODUCTION

Measures of central tendency are the statistical estimates which show the degree to which any given set of value or data will converge towards the central point of the data. It is also known as the measures of location. Basically, there are three commonest measures of central tendency namely the mean, the median and the mode. The mean shows the average value, the median shows the central number while the mode is the observation that occurs most frequent. The relationship between these three statistical variables is very important in the unit. Measures of central tendency are part of descriptive statistical analyses which aims at describing data by summarizing the values in the data set. Measures of central tendency are typical and representative of a data set. Every value in the distribution clusters around the measures of location. The population average popularly called the arithmetic mean is of such measure. An average is a value that is typical or representation of a set of data. Since such typical values tend to lie centrally within a set of data arrange according to magnitude, averages are also called measures of central tendency.

2.0 OBJECTIVES

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

(i) Explain the meaning and scope of various measures of central of tendency

(ii) Calculate the various measures of central tendency given a set of ungrouped data.

3.0 MAIN CONTENT

3.1 The Mean

This is the arithmetic average of a set of data or information. It is perhaps the most widely used measure of central tendency. It is the summation of all the total of the individual values or elements divided by making up the total.

Mathematically, arithmetic mean $(\bar{X}) = \frac{\sum fx}{\sum f}$ or $\frac{\sum_{j=1}^{N} X_j}{N}$ where the symbol $\sum_{j=1}^{N}$ is used to denote the sum of all the X_j's from j = 1 to j = N; by definition,

$$\sum_{j=1}^{N} = X_1 + X_2 + X_3 + ...X_N \text{ . E.g. } \sum_{j=1}^{6} X_j = 1+2+3+4+5+6,$$

$$\sum_{j=1}^{4} (Y_j - 3)^2 = (1-3)^2 + (2-3)^2 + (3-3)^2 + (4-4)^2$$

The mean $(\bar{X}) = \frac{x_1 + x_2 + x_3 + x_4 + \cdots x_n}{N}$

i.e. s1nt eR ntRl $(\bar{X}) = \frac{\text{um of a va ue in ttte amp e}}{\text{No of va ue in ttte amp e}}$

If x , x_2, x_3 ... x_4 have frequencies (other than 1), their frequencies are written as f, f_2, f_3 ... f_n respectively, then $(\bar{X}) = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \cdots f_nx_n}{f_1 + f_2 + f_3 + \cdots f_n}$

Then, $(\bar{X}) = \frac{\sum_{j=1}^{N} X_j}{\sum_{i=1}^{N} f_j}$ or simply $\frac{\sum fx}{\sum f}$

The major properties of arithmetic mean are:

(i) The algebraic sum of derivation of a set of number from their arithmetic mean is zero i.e. $\sum(X - \bar{X}) = 0$; where X are the individual value and $\bar{X}$ is the mean.
(ii) Every set of internal level and ratio level data has a mean.
(iii) All values are included in computing the mean.
(iv) A set of data has only one mean. It is unique.

Advantages of the Mean

(i) By making use of every value in a distribution, the mean is the most representative of all the averages (mean, median and mode).
(ii) The mean is the best known and most commonly used in statistics and management sciences.

(iii) The mean, being the result of mathematical processing, lends itself to further mathematical calculations e.g. in computing coefficient of variation, skewness, standard deviation etc.
(iv) Unlike the median and the mode, the value of the mean is always strictly determined because it is a computed average.
(v) It provides very good means of comparison.
(vi) Its value is directly affected by the value of each item in a given set of data.

Disadvantages of the Mean

(i) Extreme value in the distribution may distort the value of the mean. If the degree of such distortion is high, the mean become unrepresentative of the distribution. Thereby, becoming misleading.
(ii) If the open ended class contain a large proportion of the values, the mean may be subject to some error.
(iii) In the case of discrete data, the mean may be a value not related to any item of the distribution e.g. 8.32 workers.
(iv) Unlike the median and the mode, the mean cannot be determined graphically either by the use of frequency diagrams or cumulative frequency curves.
(v) Mean may be difficult to compute when data is numerous.

3.1.1 Mean of Ungrouped Data

Worked Examples

(1) The net weights of the content of 5 coke bottles selected at random are 85.4, 84.9, 85.3, 85.0 and 85.4. What is the arithmetic mean of the sample observation?

Solution

$$\bar{X} = \frac{\sum X}{N} = \frac{85.4 + 84.9 + 85.3 + 85.0 + 85.4}{5} = \frac{426}{5} = 85.2\text{kg}$$

(2) Given the ages of 90 students in a class as:

Ages	11	12	13	14	15
Frequency	10	20	15	25	20

Find the mean age?

Solution

Ages(X)	Frequency(F)	F(X)
11	10	110
12	20	240
13	15	195
14	25	210
15	20	300
	$\sum \mathbf{f}$ = 90	$\sum \mathbf{f}x$ = 1,055

∴ Mean ($\bar{X}$) = $\frac{\sum \mathbf{f}x}{\sum x} = \frac{055}{90} = 11.72 \quad \text{R}\mathbf{1}\mathbf{f}\text{s}$

(3) The mean score of the 3 best students in a mathematics class is 84%. The mean score of the next 4 students is 72% and the mean score of the last 5 last students is 60%. Find the mean score for the whole class of 12 students.

Mean = $\frac{\sum X}{N}$

i.e. $\bar{X}$ = $\frac{\sum X}{N}$

$\sum X = N.\bar{X}$

For the 3 best students, $\sum X \;= 84 \times 3 = 252$

For the next four students, $\sum X_2 = 72 \times 4 = 288$

For the last five students, $\sum X_3 = 60 \times 5 = 300$

∴ Total scores = 252 + 288 + 300 = 840

Total student = 3 + 4 +5 = 12

∴ Mean score of the 12 students = $\frac{\sum X_1 + \sum X_2 + \sum X_3}{N_1 + N_2 + N_3} = \frac{840}{2} = 70\%$

Weighted Arithmetic Mean

Sometimes, we associate with the numbers $x_1, x_2 \ldots x_k$, certain weighting factors (or weights): $w, w_2 \ldots w_k$, depending on the significance or importance attached to the number. In this case,

$(\bar{X}) = \frac{W_1x_1 + W_2x_2 + W_3x_3 + \cdots W_kx_k}{W_1 + W_2 + W_3 + \cdots W_k} = \frac{\sum Wk}{\sum W}$ is called the weighted arithmetic mean.

Example: The weight attached to scores in an exam is given as:

80 – 100	5 points
70 – 79	4 points
60 – 69	3 points
50 – 59	2 points
40 – 49	1 point
0 – 39	0 point

If a student sat for eight papers and have the following scores 67, 54, 92, 47, 76, 42, 31 and 82, find the weighted point average; given that the units attached to the courses are 3, 2, 1, 3, 3, 3, 2 and 2 respectively.

Solution

Scores	**Unit (F)**	**Points (X)**	**Total weight (FX)**
67	3	3	9
54	2	2	4
92	1	5	5
47	3	1	4
76	3	4	12
42	3	1	3
31	2	0	0
82	2	5	10
	$\sum \mathbf{f} = 19$		$\sum \mathbf{f}\text{x} = 47$

$\therefore$ Weighted point average = $\frac{\text{Tota Weigttt}}{\text{Tota unit}} = \frac{47}{9} = 2.47$

(2) If a final examination in a course is weighted 3 times as much as a quiz and a student has a final examination grade of 85 and quiz grades of 70 and 90, calculate the mean grade?

Solution

$$\bar{X} = \frac{70(\ )+90(\ )+85(3)}{\ +\ +3}$$

$$= \frac{4\ 5}{5} = 83$$

(3) Masco company pays its sales people ₦ 6.50, ₦ 7.50 or ₦ 8.50. The corresponding weight is 14, 10 and 2 respectively. Determine the weighted average mean.

Solution

Weighted mean $\bar{X}_W = \frac{14\times6.50+10\times7.50+2\times8.50}{14+10+2}$

$= \frac{4\ \ +75+\ \ 7}{26} = \frac{83}{26}$

$= 7.038$ = ₦ 7.04

Assumed Mean

Taking one of the values of one of the observation as the mean, the actual mean of the observation can be obtained using the formula

$$\bar{X} = A + \frac{\sum \mathbf{f}d}{\sum \mathbf{f}}$$

Where $\bar{X}$= actual mean

A = assumed mean (the value assumed to be mean)

d = deviation of each observation from the assumed mean

∴ d = X – A

Example

Given the table below, use 40 as the assumed mean to obtain the actual mean?

X	**F**
10	3
20	5
30	8
40	7
50	6
60	1

Solution: A = 40

X	**F**	**d = X – A**	**fd**
10	3	–30	–90
20	5	–20	–100
30	8	–10	–80
40	7	0	0
50	6	10	60
60	1	20	20
	$\sum \mathbf{f}$ = 30		$\sum \mathbf{f}$ = –190

$\therefore \bar{X} = A + \frac{\sum fd}{\sum f}$

$= 40 + \frac{-90}{30} = 40 - 6.33 = 33.67$

Special Means

Special means are other forms of mean apart from the arithmetic and the weighted mean. The commonest forms of special means are: -

(i) Geometric mean (G);
(ii) Harmonic mean (H); and
(iii) Quadratic mean (Q) or Root Mean Square (RMS).

The ***geometric mean*** (G) of a set of N positive numbers $X_1 . X_2 . X_3 \ldots X_n$ is the Nth root of the product of the numbers.

$\therefore G = \sqrt[N]{X_1 . X_2 . X_3 \ldots X_n}$

Example

(i) Find the geometric mean of 2, 4 and 8.

Solution

$G = \sqrt[3]{(2).(4).(8)} = \sqrt[3]{64} = 4$

(ii) What is the geometric mean of 2, 3, 4, 6, 5.

Solution

$G = \sqrt[5]{2 \times 3 \times 4 \times 6 \times 5} = \sqrt[5]{720} = 3.73$

The ***harmonic mean*** (H) of a set of N numbers $X_1 . X_2 . X_3 \ldots X_n$ is the reciprocal of the arithmetic mean of the reciprocals of the numbers.

$$\therefore H = \frac{1}{\frac{1}{N}\sum_{j=1}^{N}\frac{1}{X_j}} = \frac{N}{\sum \frac{1}{X}} \text{ or } N.\frac{1}{\sum \frac{1}{X}}$$

Example

Find the harmonic mean of (a) 2, 4, 8 and (b) 2, 3, 4, 5, 6.

Solution

(a) $H = \frac{N}{\sum \frac{1}{X}} = \frac{3}{\frac{1}{2}+\frac{1}{4}+\frac{1}{8}} = \frac{3}{\frac{7}{8}} = \frac{3\times 8}{7} = \frac{24}{7} = 3.43$

(b) $H = \frac{N}{\sum \frac{1}{X}} = \frac{5}{\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}} = \frac{5}{\frac{87}{60}} = \frac{5\times 60}{87} = \frac{300}{87} = 3.45$

The ***root mean square*** (RMS) or the quadratic mean of a set of numbers $X_1 . X_2 . X_3 \ldots X_n$ is sometimes denoted by $\sqrt{\bar{X^2}}$ of $\sqrt{\frac{\sum X^2}{N}}$. It is defined as the square root of the arithmetic mean of their squares. It is frequently used in physical applications.

Example

Find the RMS of the following seta of data

(a) 1, 3, 4, 5, 7
(b) 1, 2, 4, 5, 3

Solution

(a) RMS = $\sqrt{\frac{{}^2+3^2+4^2+5^2+7^2}{5}} = \sqrt{\frac{00}{5}} = \sqrt{20} = 4.47$

(b) RMS = $\sqrt{\frac{{}^2+2^2+4^2+5^2+3^2}{5}} = \sqrt{\frac{90}{5}} = \sqrt{16} = 4$

The relationship between arithmetic geometric and harmonic means is that the geometric mean of a set of number $X_1, X_2, X_3 \ldots X_n$ is less than or equal to their arithmetic mean but greater or equal to their harmonic mean. In symbols, $H \le G \le \bar{X}$. The equality signs hold only if all the numbers $X_1 . X_2 . X_3 \ldots X_n$ are identical. For example, for the set of data: 2, 4, 8, the arithmetic mean is 4.57, geometric mean is 4 and the harmonic mean is 3.43.

$$3.43 \le 4 \le 4.67$$

i.e. $H \le G \le \bar{X}$

CLASS ASSESSMENT EXERCISE 1

(1) Write out the terms in each of the followings:

(i) $\sum_{j=1}^{8} X_j$ (ii) $\sum_{j=1}^{N} a$ (ii) $\sum_{j=1}^{3} (X_j - a)$ (iv) $\sum_{i=1}^{5} (X_i + 2)$

(2) The data below shows the size of shoes of 10 set of students: 3, 5, 6, 6, 7, 10, 10, 16, 18 and 20. Calculate:
(i) Arithmetic mean
(ii) Harmonic mean
(iii) Geometric mean
(iv) Quadratic mean

(3) Given the scores of some students in a test as

Scores	0	1	2	3	4	5	6	7	8
Frequency	1	8	28	56	70	56	28	8	1

(a) Find the mean mark?
(b) Use assume mean of 5 to estimate the mean mark.

3.2 The Median

When data consisting of n-members is arranged in order of magnitude, the middle value or member is called the ***median*** of the data if *n* is odd but the arithmetic mean of the middle value is the median if *n* is even. It is the value which divides the distribution into two so that an equal number of the value lies below and above it.

Generally, the median of a set of n-numbers is defined as the ½(n + 1)th item or value where the n-number are arranged in ascending or descending order of magnitude.

The properties of the median include: -

(i) It is unique i.e. like the mean, there is only one median for set of dat.
(ii) To determine the median, arrange the data from the lowest to the highest or otherwise hence, find the value of the middle observation.
(iii) The median is not affected by the extreme large or small values; therefore, it (the median) is a valuable measure of central tendency.

Advantages of the Median

(i) It is easy to calculate and understand.
(ii) It depends on the middle items or groups; it is not affected by the extreme values.
(iii) It can be calculated from incomplete data.
(iv) It is an actual value occurring in the distribution and therefore, it is related to the value in the distribution.
(v) The median can be determined from frequency diagram i.e. it can be obtained graphically.
(vi) It gives a clear idea of the distribution of the data.

Disadvantages of the Median

(i) It is not useful in further statistical calculations.
(ii) It may require re-arrangement of data involved before calculation.
(iii) It is difficult to use with discrete variable which takes only value.
(iv) It takes no account of extreme values in the observation.
(v) Irregular distribution makes the location of the median to be indefinite.
(vi) It is a position average, it has limited practical application.
(vii) It is unrepresentative of all the values in the distribution and to make further use of it, it must be supplemented by other statistics.

Worked Examples

1. Find the median of each of the following sets of information
   (a) 3, 2, 2, 5, 1, 4, 3, 2, 1, 5 and 2.
   (b) 3, 6, 5, 4, 2, 4, 8, 4, 6, 8, 9 and 10.

Solution

(a) Arrange the data

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

There are 11 numbers in the set. The median is the 6the item.

i.e. ½(11 + 1)th item = ½(12)th item = 6th item

(b) Arranging the data:

2, 3, 4, 4, 5, 6, 6, 8, 8, 9, 9, 10

The median is ½ (12 + 1)th item = ½ (13th item)

$= \frac{3}{2} ah\ taRnt = 6\frac{\ }{2} ah\ taRnt$

$= \frac{6ttt\ item + 7ttt\ \ item}{2} = \frac{6+6}{2} = 6$

2. Given the score of 120 students in an examination as:

Scores	0	1	2	3	4	5	6	7	8	9	10
Frequency	7	10	13	10	15	5	18	12	5	10	15

Find the median?

Solution

It may be tedious writing out the scores in an array in order of magnitude to locate the middle value(s). It is advisable to draw a cumulative frequency table to address this.

Scores	**Frequency (f)**	**Cumulative frequency**
0	7	7
1	10	17
2	13	30
3	10	40
4	15	55
5	5	60
6	18	78
7	12	90
8	5	95
9	10	105
10	15	120
	$\sum \mathbf{f} = N = 120$	

Median $= \frac{\ }{2}(N + 1) ah\ \ taRnt$

$= \frac{\ }{2}(120 + 1) ah\ \ taRnt$

$= \frac{2}{2}\ \ ah\ taRnt$

$= 60.5 ah\ taRnt$

$$= \frac{60\text{ttt item} + 6 \quad \text{t item}}{2}$$

From the cumulative frequency distribution, the items in this position are 5 and 6. Therefore, the median = $= \frac{5+6}{2} = 5.5$

CLASS ASSESSMENT EXERCISE 2

1. Find the median of the following set of data
   (a) 7, 8, 6, 5, 4, 9, 2, 6, 8, 1, 5, 7, 2, 9.
   (b) 6, 7, 5, 6, 4, 2, 1, 8, 10, 16, 13, 29, 23.
2. Given the ages of 400 pensioners as:

Ages	60	61	62	63	64	65	66	67	68	69	70
Freq.	50	42	54	21	40	48	47	25	28	13	32

   Find the median age?

3.3 The Mode

The mode is the variable occurring most which corresponds to the highest point of the frequency curve. It is the value or item that has the highest frequency. An observation having one mode is said to be ***Unimodal***. Those observation having two modes are said to be ***Bimodal*** while those having more than two modes are ***Multimodal***.

Advantages of Mode

(i) It is easy to obtain either graphically or manually.
(ii) It is not be affected by extreme values.
(iii) It is quickly obtained, realistic and dependable.
(iv) It is a good representation of the data.
(v) It is not affected by open-ended classes or extreme values of the distribution.

Disadvantages of Mode

(i) It is not useful for further mathematical management.
(ii) It is not a very good measure of central tendency.
(iii) It is difficult to obtain in a large and grouped data.
(iv) It has limited practical use in management.
(v) It is not necessarily unique as there can be more than one mode in a set of data.
(vi) It does not represent all the values in the distribution.

Mode of Ungrouped Data

For ungrouped data, the mode can be obtained by inspection e.g.

(i) Given the scores of 10 students as: 5, 4, 6, 2, 7, 4, 9, 2, 4. What is the mode?

Solution

The mode = 4 (the most frequent occurring score).

(ii) Given the frequency distribution as thus:

Scores	10	11	12	13	14	15	16	17	1
Freq.	2	5	1	2	1	1	5	2	4

What is the modal score?

Solution

The modes are 11 and 16 (because each of them has the highest frequencies of 5).

4.0 CONCLUSION

This unit has been able to introduce the students to the basic measure of central tendency of ungrouped data. Their definition properties advantages, disadvantages and computation.

5.0 SUMMARY

Mean, median and mode are the basic measures of central tendency or the measures of location. While the mean is the arithmetic average, the median is the locational average or the middle-value and the mode is the most occurring item in an observation. For a unimodal frequency curve that are moderately skewed (asymmetrical), the relationship between mean, median and mode is given as Mean – Mode = 3 (Mean – Median)

6.0 TUTOR MARKED ASSIGNMENT

1. The table below shows the marks scored by 32 students in a test.

Mark Scored	1	2	3	4	5	6	7	8	9	10
No of Students	2	3	4	4	4	4	5	3	2	1

What is the modal mark? What type of mode is this?

7.0 REFERENCES/FURTHER READING

Akanbi, S. O. and Chika, N. C. (2007). Mathematics for Senior Secondary School. Macmillan Publishing, Nigeria.

Channon, J. B. et al (2002). New General Mathematics for Senior Secondary Scholls 3, Longman Publishers, Nigeria.

Spiegel, M. R. and Stephens, L. T. (2009). Statistics. Schaums Outline Series (4th Edition), McGraw Hills Publishers, Delhi.

UNIT 2

MEASURES OF CENTRAL TENDENCY OF GROUPED DATA

Table of Contents

1.0 INTRODUCTION

In the previous unit, the measures of central tendency have been considered for ungrouped data. This unit attempts at studying the basic approaches involved in computing the measures of central tendency of grouped data as well as deriving them from curves and histograms.

2.0 OBJECTIVES

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

(i) Compute the measures of central tendency for grouped data;
(ii) Estimate some measures of central tendency from curves and histogram

3.0 MAIN CONTENT

3.1 Mean of Grouped Data

It should be noted that in grouped frequency distribution, the values between any class intervals are considered as condensed or concentrated at the midpoint of the class interval or the class mark.

So, if x_i is the class mark of the ith class interval, then the mean ($\bar{X}$) of the grouped frequency distribution is also defined by:

$$\bar{X} = \frac{\sum_{i=1}^{N} f_i X_i}{\sum_{i=1}^{N} f_i} = \frac{\sum fx}{\sum f}$$

Examples

(1) Given that the information below

Height(inches)	60 – 62	63 – 65	66 – 68	69 – 71	72 – 74
Frequency	5	18	42	27	8

Find the mean mark?

(2) The following table shows the distribution of weekly wages earned by a number of employees of FOC Construction Company in Lagos.

Wages (₦)	**Number of employees**
40 – 49	4
50 – 59	12
60 – 69	18
70 – 79	11
80 – 89	7
90 – 99	5
100 – 109	2
110 – 119	1

Using an assumed mean of ₦ 74.50, calculate the mean of the distribution?

Solution

(1)

Height(inches)	**Freq(f)**	**Midpoint (x)**	**(fx)**
60 – 62	5	61	305
63 – 65	18	64	1152
66 – 68	42	67	2814
69 – 71	27	70	1890
72 – 74	8	73	584
	$\sum \mathbf{f} = 100$		$\sum \mathbf{f}x = 6745$

∴ Mean = $\frac{\sum \mathbf{f}x}{\sum \mathbf{f}} = \frac{6745}{00} = 6745$

(2)

Wages (₦)	Class mark	F	d = X – A	fd
40 – 49	44.5	4	–30	–120
50 – 59	54.5	12	–20	–240
60 – 69	64.5	18	–10	–180
70 – 79	74.5	11	0	0
80 – 89	84.5	7	10	70
90 – 99	94.5	5	20	100
100 – 109	104.5	2	30	60
110 – 119	114.5	1	40	40
		60		$\sum f$ = –270

$\bar{X} = A + \frac{\sum fd}{\sum f}$

$\bar{X} = 74.5 + \ -\frac{270}{60}$

$\bar{X} = 74.5 - 4.5$

$\bar{X} =$ ₦ 70.00

CLASS ASSESSMENT EXERCISE 1

1. Given the ages of 60 pupils in a class as follows: -

18	20	20	21	21	22	23	23	20
10	11	12	11	10	13	12	12	10
14	11	18	16	14	16	13	13	11
11	14	10	13	14	16	18	18	14
10	13	12	15	14	16	14	14	13
19	27	18	21	23	24	28	28	27

Prepare the frequency distribution table using the class size 10 – 13, 14 – 17, 18 – 21, 22 – 25; find:

(a) The mean age
(b) The actual mean age, using the assumed mean of 20 years
(c) Compare your answer in (a) and (b) above.

3.2 Median of Grouped Data

Given a set of grouped data, the median is given by:

$$\text{Median} = L + \frac{\frac{N}{2} - m \; c}{f}$$

Where L = lower limit or lower class boundary of the median class.

N = Number of items or $\sum f$

m = cumulative frequency just before the median class.

f = frequency of the median class

c = class limit

Example

Given age distribution of a set of people as:

Age Group	1 – 20	21 – 40	41 – 60	61 – 80	81 – 100	101 – 120
Frequency	2	2	14	14	16	2

Find the median age?

Solution

Classes	Frequency	Cumm. Freq.	Class Boundaries
1 – 20	2	2	0.5 – 20.5
21 – 40	2	4	20.5 – 40.5
41 – 60	14	18	40.5 – 60.5
61 – 80	14	32	60.5 – 80.5
81 – 100	16	48	80.5 – 100.5
101 – 120	2	50	100.5 – 120.5

$$\text{Median} = L + \frac{\frac{N}{2} - m \; c}{f}$$

$$= 60.5 + \frac{\frac{50}{2} - 8 \; 20}{4}$$

$$= 60.5 + \frac{40}{4}$$

$$= 60.5 + 10$$

$= 70.5$ Rlfs

Geometric Determination of the Median

The median of a set of grouped data can be determined geometrically in two ways:

(a) From a histogram
(b) From a cumulative frequency curve

Median from a Histogram: - The median determined from a histogram, is the value on the variable axis through which a vertical line, dividing the histogram into two equal areas, passes.

Median from Cumulative Frequency Curve: - The median can be determined from a cumulative frequency curve as follows. Through the point which divides the cumulative frequency axis into two equal parts, draw a line parallel to the variable axis to intersect the cumulative frequency curve. From the point of intersection on the cumulative frequency curve, drop a perpendicular to the variable axis. The point where this vertical line meets the variable axis is the median of the distribution.

Note: For a grouped frequency distribution, the class interval which contains the median value is the median class.

Example

The following table shows the distribution of the masses of 120 logs of wood correct to the nearest kg.

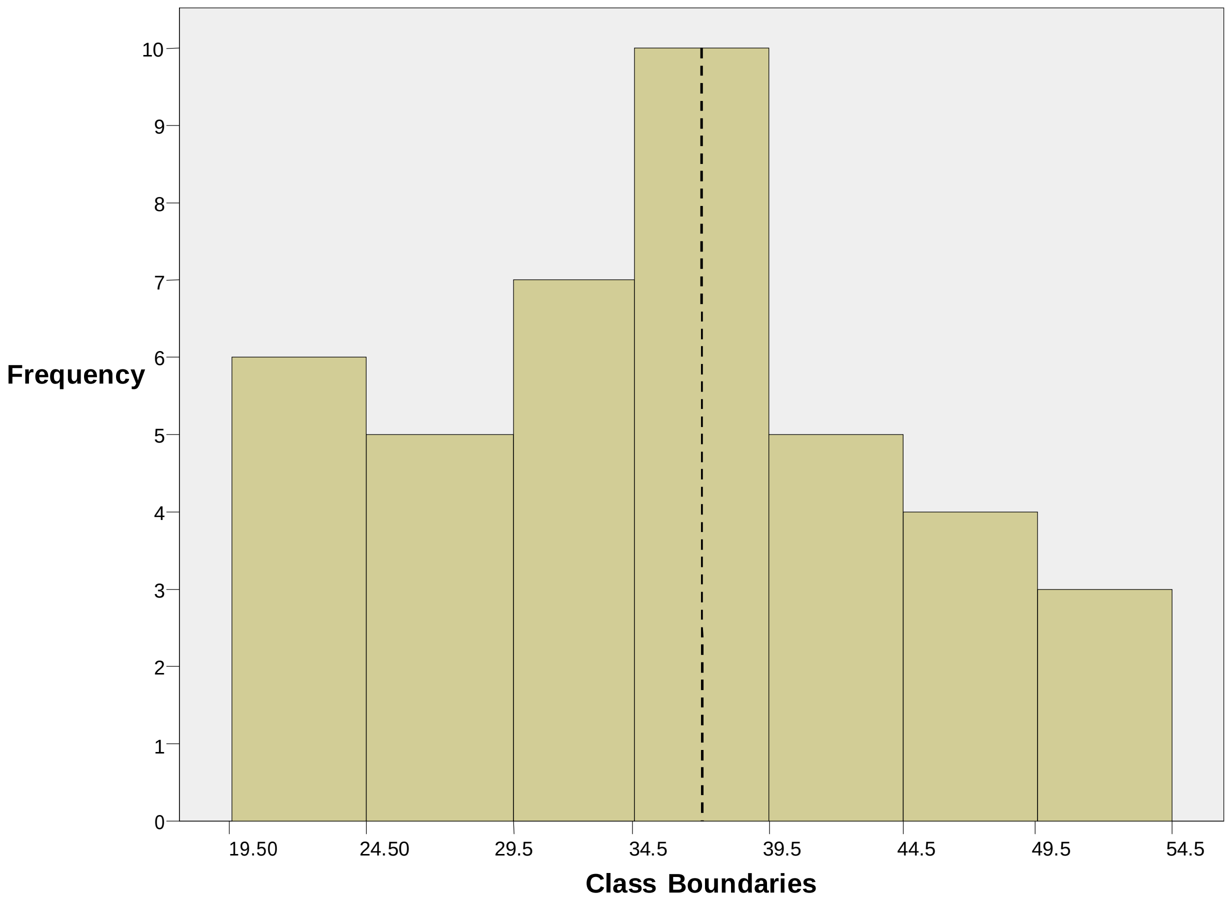
Mass/kg	Frequency	Class Boundaries	Cumm. Freq.
20 – 24	6	19.5 – 24.5	6
25 – 29	5	24.5 – 29.5	11
30 – 34	7	29.5 – 34.5	18
35 – 39	10	34.5 – 39.5	28
40 – 44	5	39.5 – 44.5	33
45 – 49	4	44.5 – 49.5	37
50 – 54	3	49.5 – 54.5	40

Required: (a) Present the detail with the aid of histogram and use it to estimate the median.

(b) Present the data with the aid of cumulative frequency curve and use it to estimate the median.

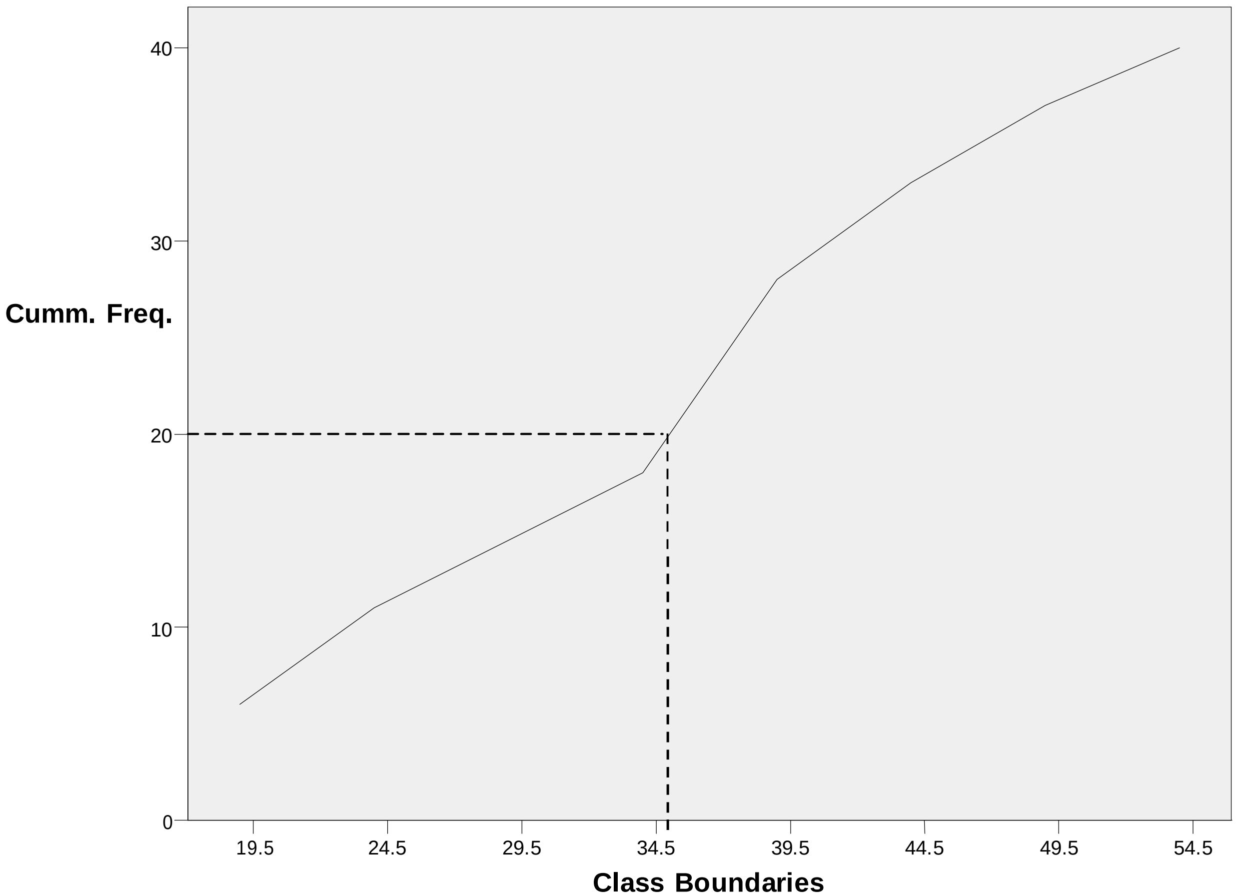
Solution

(a) Histogram Showing Distribution of Masses of Logs of Woods



The median is approximately 36

b) Cumulative frequency Curve showing the distribution of masses of logs of wood



Median is approximately 35.5

CLASS ASSESSMENT EXERCISE 2

1. (a) Find the median of the information below:

Scores	1 – 20	21 – 40	41 – 60	61 – 80	81 – 100
Frequency	7	12	15	19	7

(b) Present the data using histogram.
(c) Present the data using cumulative frequency curve.
(d) Estimate the median from (b) and (c) above.

3.3 Mode of Grouped Data

There are 2 common methods or formulae of obtaining the mode of grouped data: -

(i) $\text{Mode} = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \, c$

Where;

L = lower class boundary of the class that has the modal frequency (modal frequency class).

f_0 = frequency of the grouped/class before the modal frequency class.

f_1 = frequency of the group that contains the mode.

f_2 = frequency of the group after the modal class group.

c = class interval

(ii) $\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \, c$

Where;

Δ_1 = difference between the frequency of modal class and the frequency of the class before it.

Δ_2 = difference between the frequency of the modal class and the frequency of the class after it.

Example

Given the data below:

Class Interval	20 – 29	30 – 39	40 – 49	50 – 59	60 – 69	70 – 79	80 – 89	90 – 99
Freq.	9	10	13	19	11	8	6	4

Calculate the mode using each of the formula stated above.

(i) $\text{Mode} = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \, c$

L = 49.5

$f_0 = 13$

$f_1 = 19$

$f_2 = 11$

c = 10

$\text{Mode} = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \, c$

$$= 49.5 + \frac{19 - 13}{2(19) - 13 - 11} \, 10$$

$$= 49.5 + \frac{6}{38 - 13 - 11} \, 10$$

$$= 49.5 + (4.28) = 53.78$$

(ii) Mode $= L + \frac{\Delta_1}{\Delta_1 + \Delta_2} C$

$\Delta = 19 - 13 = 6$

$\Delta_2 = 19 - 11 = 8$

$= 49.5 + \frac{6}{6+8}\ 10$

$= 49.5 + \frac{60}{4}$

$= 49.5 + 4.28 = 53.78$

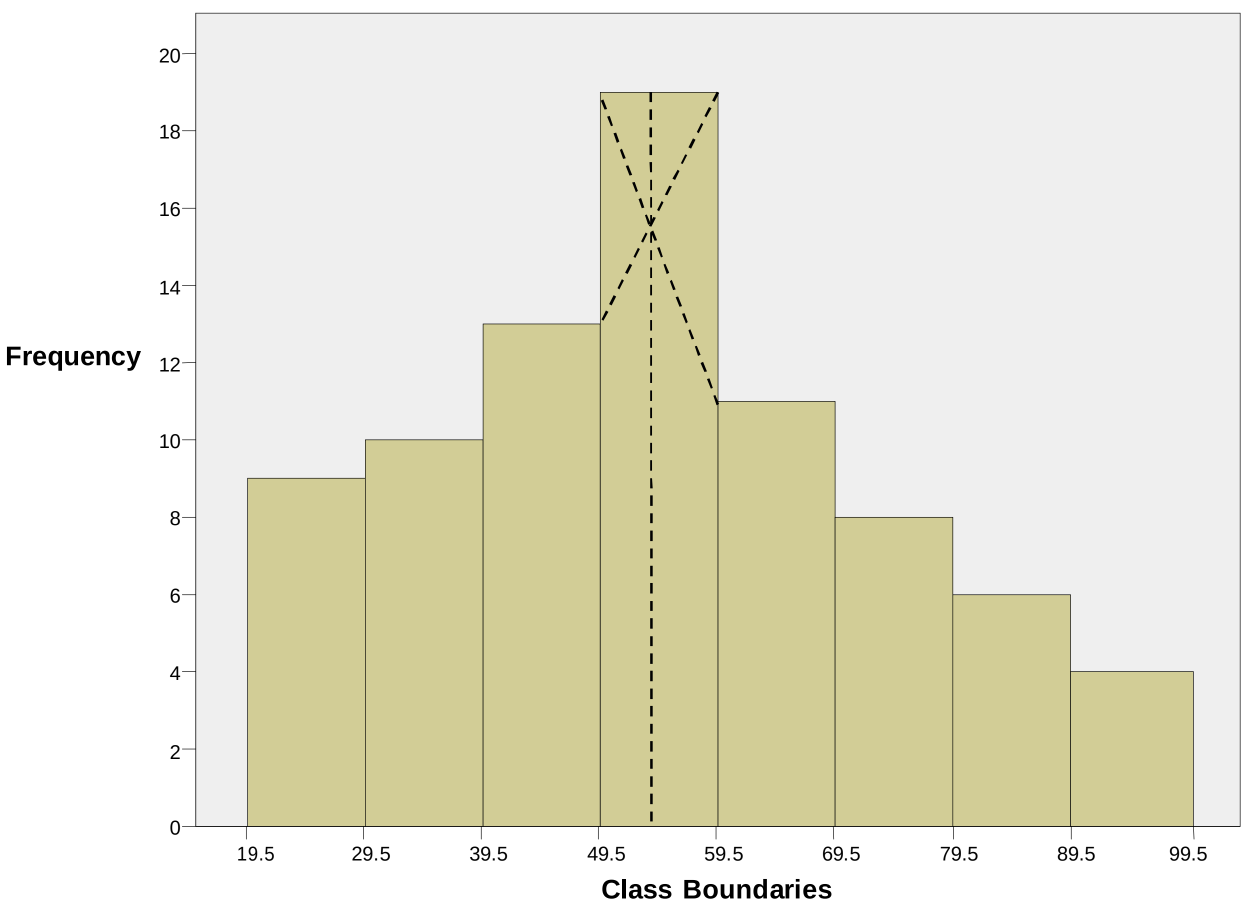
Geometric Determination of the Mode: - The mode of a grouped data can be obtained from the histogram. This can be done by drawing a straight line from the opposite extreme end of the bars beside the tallest bar. The straight lines should be drawn across the highest bar. The point of intersection of the straight line is traced to the horizontal or variable axis to give the mode.

Example

Use the histogram to find the mode of the data below:

Scores	20 – 29	30 – 39	40 – 49	50 – 59	60 – 69	70 – 79	80 – 89	90 – 99
Freq.	9	10	13	19	11	8	6	4

Histogram of Scores



Mode is approximately = 54

CLASS ASSESSMENT EXERCISE

1. Given the scores of students in an examination as:

Score	0 – 9	10 – 19	20 – 29	30 – 39	40 – 49	50 – 59	60 – 69	70 – 79	80 – 89
Freq.	2	2	14	14	16	2	15	25	10

(a) Find the mode?
(b) Use histogram to estimate the mode.

4.0 CONCLUSION

This unit completes section on measures of central tendency by paying specific attention to the grouped data.

5.0 SUMMARY

Due to the specific nature of grouped data i.e. more than a variable forming a class, there is need to derive means of computing the measure of central tendency with the use of peculiar features of grouped data such as the midpoint, class interval, class boundaries etc. Specifically, some measure of central tendency of grouped data can also be obtained directly from charts and curves such as the histogram and cumulative frequency curves.

6.0 TUTOR MARKED ASSIGNMENT

1. Below is a table showing the marks obtained by a sample of 200 students in economics at the end of a session.

Marked Obtained	Frequency
Under 30	2
30 – 39	4
40 – 49	12
50 – 59	50
60 – 69	73
70 – 79	43
80 – 89	13
90 – 99	3

(a) Find the mean mark, the median mark and the modal mark.
(b) Derive the median and the mode geometrically.
(c) Compare your results in (b) to the answers obtained in (a)

7.0 REFERENCES/FURTHER READING

Akanbi, S. O. and Chika, N. C. (2007). Mathematics for Senior Secondary School. Macmillan Publishing, Nigeria.

Channon, J. B. et al (2002). New General Mathematics for Senior Secondary Scholls 3, Longman Publishers, Nigeria.

Spiegel, M. R. and Stephens, L. T. (2009). Statistics. Schaums Outline Series (4th Edition), McGraw Hills Publishers, Delhi.

UNIT 3

MEASURES OF DISPERSION

Table of Contents	

1.0 INTRODUCTION

The measure of dispersion estimate the extent or degree to which values in a set of data tend to spread around or about the average value. The commonest measures of dispersion are:

(i) The Range

(ii) The Mean Deviation

(iii) The Standard Deviation

(iv) The Variance

2.0 OBJECTIVES

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

(i) Define the range, mean deviation, standard deviation and variance.

(ii) Compute and interpret the range, men deviation, standard deviation and variance of difference for difference forms of data.

3.0 MAIN CONTENT

3.1 The Range

This is the simplest of all the measure of dispersion. It is the difference between the highest and the lowest value in a set of data. Mathematically, the range = highest observation – lowest observation. However, for grouped data, the range is the difference between the highest upper class boundary and the lowest class boundary. The higher the value of the range, the greater the absolute variation in the extreme values of the distribution.

Advantages of Range

(i) It is useful for further statistical calculation.
(ii) It gives a rough estimate of the difference between the values to be handled.
(iii) It is easy to calculate and understand.
(iv) It helps in keeping variability in check.

Disadvantage of Range

(i) It does not consider all the values in the data.
(ii) It is not a reliable measure of variability because at times, two different data may have the same range even though, their dispersion may be different.

Note: - Relative Range = $\frac{X_{max} - X_{min}}{X_{min} + X_{max}}$

Where; X_{max} = Highest value and X_{min} = Lowest value

Mid Range = $\frac{\text{Range}}{2}$

Example

The scores of 20students in a mathematics test are as follows:
27, 16, 10, 19, 10, 11, 28, 16, 14, 35
3, 4, 9, 22, 21, 26, 27, 40, 41, 3. Find the range.
Solution

The highest score = 41
The least/lowest score = 3
∴ The range = Highest Score – Lowest Score
= 41 – 3 = 38
CLASS ASSESSMENT EXERCISE 1

(a) Find the range and the relative of the following set of ages: 2, 3, 1, 2, 4, 3, 2, 3, 2 and 1.
(b) What is the mid range?

3.2 Mean Deviation

The mean deviation refers to the arithmetic average of all deviation in a distribution from the mean. The mean deviation or average deviation, of a set of *N* numbers: X_1, X_2 ... X_N is abbreviated as *MD* and is defined by:

MD = $\frac{\sum|X-\bar{X}|}{N}$, where $\bar{X}$ is the arithmetic mean of the numbers and $|X - \bar{X}|$ is the absolute value of the deviation X_j from $\bar{X}$. The absolute value of number is the number without the associated sign and is indicated by two vertical lines placed around the number; thus, $|-4| = 4, |+3| = 3, |6| = 6$ **1** $|-0.84| = 0.84$).

Occasionally, the mean deviation is defined in terms of absolute deviations from the median or other average instead from the mean. An interesting property of the sum $\sum_{j=1}^{N} |Xj - a|$ is that it is a minimum where *a* is the median (i.e. the mean deviation about the median is minimum). Note that it would be more appropriate to use the terminology mean absolute deviation than men deviation.

The sum of all deviation from the mean is always zero (0). A lower mean deviation implies less dispersion while a higher one implies large dispersion. For grouped data or discrete data with frequency (f), the mean deviation

MD = $\frac{\sum f|X-\bar{X}|}{\sum f}$ where; $|X - \bar{X}| =$ (the deviation from the mean),

Then, MD = $\frac{\sum f|d|}{\sum f}$

One advantage of mean of mean deviation is that, it presents a good picture of the data because every item is taken into account. However, the major disadvantages or shortcomings of mean deviation are: -

(i) It is not useful for further mathematical analysis.
(ii) It may be difficult to calculate when observation is numerous.

Worked Examples

1. Find the mean deviation of the data given below:

Scores(X)	0	1	2	3	4	5	6	7	8
Frequency(f)	1	8	28	56	70	56	28	8	1

2. Given the information below as the age distribution members of a family.

Age Limits	**Frequency**
1 – 20	1
21 – 40	2
41 – 60	14
61 – 80	14
81 – 100	16
101 – 120	3

Required: Find the mean deviation?

Solution

1.

X	f	f(X)	$(\;\; - \bar{\;})$	$\| \;\; - \bar{\;} \|$	$t\| \;\; - \bar{\;}\|$
0	1	0	-4	4	4
1	8	8	-3	3	24
2	28	56	-2	2	56
3	56	168	-1	1	56
4	70	240	0	0	0
5	56	280	1	1	56
6	28	168	2	2	56
7	8	56	3	3	24
8	1	8	4	4	4
	$\sum \mathbf{f} = 256$	$\sum \mathbf{f}X = 1024$	$\sum(X - \bar{X}) = 0$	$\sum\|X - \bar{X}\| = 20$	$\sum \mathbf{f}\|X - \bar{X}\| = 280$

Mean $(\bar{\;}) = \frac{\sum tx}{\sum t} = \frac{1o24}{2s6} = 4$

Mean Deviation = $\frac{\sum \mathbf{f}|X-\bar{X}|}{N} = \frac{280}{256} = 1.093$

2.

Class Limit	Freq(f)	Midpoint	f(X)	$(\;\; - \bar{\;})$	$\| \;\; - \bar{\;}\|$	$t\| \;\; - \bar{\;}\|$
1 – 20	1	10.5	10.5	-60.4	60.4	60.4
21 – 40	2	30.5	61	-40.4	40.4	80.5
41 – 60	14	50.5	707	-20.4	20.4	285.6
61 – 80	14	70.5	987	0.4	0.4	5.6
81 – 100	16	90.5	1448	19.6	19.6	313.6
101 – 120	3	110.5	331.5	39.6	39.6	118.8
	$\sum \mathbf{f} = 50$		$\sum \mathbf{f}X =$ 3545			$\sum \mathbf{f}\|X - \bar{X}\|$ = 864.8

Mean $(\bar{\;}) = \frac{\sum tx}{\sum t} = \frac{3s4s}{so} = 70.9$

Mean Deviation = $\frac{\sum \mathbf{f}|X-\bar{X}|}{N} = \frac{864.8}{50} = 17.296$

CLASS ASSESSMENT EXERCISE 2

(1) The table below gives the number of hours, that is factory workers spent on a job

No of Hours(X)	2	3	4	5	6	7	9	12
No of Workers(f)	1	2	1	3	3	2	2	1

Calculate the mean deviation of the information.

(2) The table below gives the degree of workers lateness to work in a day.

Period of lateness (in minutes)	Frequency
1 – 25	10
6 – 10	7
11 – 15	2
16 – 20	7
21 – 25	4

Find the men deviation.

3.3 Standard Deviation and Variance

All the measures of dispersion which we have discussed so far have one or two major setbacks which make their reliability questionable. The range makes use of two extreme observation s and it is possible for two sets of data to have the same range but different degree of spread about the measure of central tendency. The mean deviation, although more reliable than the range, by taking into account all the observation. However, its own setback is that, the deviation of an observation from a measure of central tendency has to be made positive, even when it is negative. The making of the deviations positive when they are negative requires extra mathematical justifications. For this reason, the mean deviation may not be an ideal measure of dispersion.

Hence, one measure of dispersion which is very reliable is the variance of the mean square deviation. It takes the arithmetic mean of the squares of deviation of the observation from the true mean. The positive square root of the mean squared deviation is called the standard deviation. As the name suggests, mathematicians and statisticians have agreed that the standard deviation is the most reliable measure of dispersion. It is usually denoted *S* or a and has the same unit as those of the observation. The square of the standard deviation is the variance denoted by S^2 or a^2 as the square root of variance is the standard deviation.

By definition, standard deviation is the square root of the arithmetic mean of the sum of squares of deviation of the values in the distribution from the mean. When the observations are clustered around the mean, the standard deviation is small but when they spread out, the standard deviation will be large.

For ungrouped data, standard deviation

$$= \sqrt{\frac{\sum(X-\bar{X})^2}{n}} \text{ of } \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2}$$

Where; X = each observation and $\bar{X}$ = mean.

For grouped data, standard deviation

$$= \sqrt{\frac{\sum f(X-\bar{X})^2}{\sum f}} \text{ of } \sqrt{\frac{\sum fX^2}{\sum f} - \left(\frac{\sum fX}{\sum f}\right)^2}$$

Where; X = midpoint and $\bar{X}$ = mean.

Given that, A = assumed mean, c = class interval and X = midpoint; $U = \frac{x-A}{c}$

$\therefore \quad = C\sqrt{\frac{\sum fu^2}{\sum f} - \left(\frac{\sum fu}{\sum f}\right)^2}$. This is called coding formula for finding the standard deviation of grouped data. It is often used for frequency distribution of equal class interval.

Variance is the square of standard deviation. For ungrouped data,

$$S^2 \text{ or } a^2 = \frac{\sum(X-\bar{X})^2}{N-} \text{ of } \frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2$$

While for grouped data, the variance is given as:

$$S^2 \text{ or } a^2 = \frac{\sum f(X-\bar{X})^2}{N-} \text{ of } \frac{\sum fX^2}{\sum f} - \left(\frac{\sum fX}{\sum f}\right)^2$$

Where; X = midpoint and $\bar{X}$ = mean.

Using the coding formula, standard deviation

$$S^2 \text{ or } a^2 = C\left(\frac{\sum fu^2}{\sum f} - \left(\frac{\sum fu}{\sum f}\right)^2\right)$$

Note: -

(i) For moderately skewed distribution, we have empirical relationship as

Mean deviation = $\frac{4}{5}$(standard deviation)

Semi-inter quartile range = $\frac{2}{3}$(standard deviation)

Where semi-inter quartile range = $\frac{Q_3-Q_1}{2}$

Q_3 = upper quartile

Q_1 = lower quartile (details about quartiles shall be discussed in the next unit)

(ii) The actual variation or dispersion, as determined from the standard deviation or other measure of dispersion is called the Absolute dispersion. However, a variation (or dispersion) of 10 inches (inch) in measuring a distance of 1000 feet (ft) is quite different in effect from the same variation of 10 inches in a distance of 20 ft. A measure of this effect is supplied by the relative dispersion, which is defined by:

Relative dispersion = $\frac{\text{ab o ute di per ion}}{\text{average}}$

If the absolute dispersion is the standard deviation S, and if the average is the mean $\bar{X}$, then the relative dispersion is called the co-efficient of variation or co-efficient of dispersion, denoted by V, given as:

Co-efficient of variation (V) = $\frac{S}{\bar{X}}$ (usually expressed as a percentage).

It should be noted that, the co-efficient of variation is independent of the units used. For this reason, it is useful in comparing distributions where the units may be different. A disadvantage of the co-efficient of variation is that, it fails to be useful when $\bar{X}$ is close to zero.

Worked Examples

1. Given the scores of 8 students in an examination as 12, 5, 6, 2, 14, 10, 18 and 5. Find the standard deviation, variance and the co-efficient of variation?
2. The following table shows the distribution wage earned by a number of employees of F.O.C. Construction Company in Lagos.

Wages (₦)	40–4 9	50–59	60–69	70–79	80–89	90–99	100–109	110–119
No of employees	4	12	18	11	7	5	2	1

(a) Find the standard deviation and variance.
(b) Find the standard deviation and variance using assumed mean of ₦ 74.50.
(c) Determine the co-efficient of variation from (b) above.

Solution

1.

Scores (X)	**(− ¯)**	**(− ¯)^2**
12	3	9
5	-4	16
6	-3	9
2	-7	49
14	5	25
10	1	1
18	9	81
5	-4	16
		∑(− ¯) = 206

Mean = $\frac{\Sigma}{N} = \frac{2+5+6+2+\ 4+\ 0+\ 8+5}{8}$

$= \frac{72}{8} = 9$

∴ Standard deviation = $S = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n}} = \sqrt{\frac{206}{8}} = \sqrt{25.75} = 5.07$

Variance = $S^2 = (5.07)^2 = 25.75$

Co-efficient of variation = $\frac{S}{\bar{X}} \times \frac{00}{}$

$$= \frac{5.07}{9} \times \frac{00}{} = 56.33\%$$

2a.

Wages	Freq(f)	Midpoint	f(X)	(− ⁻)	(− ⁻)^2	ƭ(− ⁻)^2	X^2	fX^2
40–49	4	44.5	178	-25.5	650.25	2601	1980.25	7921
50–59	12	54.5	654	-15.5	240.25	2883	2970.25	35643
60–69	18	64.5	1161	-5.5	0.25	544.5	4160.25	74884.5
70–79	11	74.5	819.5	4.5	20.25	222.75	5550.25	61052.75
80–89	7	84.5	591.5	14.5	210.25	1471.75	7140.25	49981.75
90–99	5	94.5	472.5	24.5	600.25	300`.25	8930.25	44651.25
100–109	2	100.5	209	34.5	1190.25	2380.5	10920.25	21840.5
110–119	1	105.5	114.5	44.5	1980.25	114.5	13710.25	13110.25
	$\sum f$ = 60		$\sum fX$ = 4200			$\sum f(X - \bar{X})$ = 13219.25		$\sum fX^2$=3 09085

Mean $\bar{X} = \frac{\sum fX}{\sum f} = \frac{4200}{60} = 70$

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

$$= \sqrt{\frac{309085}{60} - \left(\frac{4200}{60}\right)^2}$$
$$= \sqrt{5151.4 - (70)^2}$$
$$= \sqrt{5151.4 - 4900}$$
$$= \sqrt{251.4}$$
$$= 15.86$$

b.

Wages	Class mark	f	d = x–A	$u = \frac{x - A}{c}$	u^2	f(u)	f(u^2)
40–49	44.5	4	-30	-3	9	-12	36
50–59	54.5	12	-20	-2	4	-24	48
60–69	64.5	18	-10	-1	1	-18	18
70–79	74.5	11	0	0	0	0	0
80–89	84.5	7	10	1	1	7	7
90–99	94.5	5	20	2	4	10	20
100–109	100.5	2	30	3	9	6	18
110–119	105.5	1	40	4	16	4	16
		$\sum f$ = 60				$\sum f q$ = -27	$\sum f q^2$= 163

A = 74.5, c = 10, $\sum \mathbf{f} = \mathbf{60}$**,** $\sum \mathbf{f}\, q = 27$**,** $\sum \mathbf{f}\, q^2 = 163$

$$X = A + C \ \frac{\sum \mathbf{fu}}{\sum \mathbf{f}}$$

$$= 74.5 + 10 \ \ -\frac{27}{60}$$

$$= 74.5 - 4.5 = 70$$

$$\therefore \ \ = C \ \sqrt{\frac{\sum \mathbf{fu}^2}{\sum \mathbf{f}} - \ \frac{\sum \mathbf{fu}}{\sum \mathbf{f}}^{\ 2}}$$

$$= 10 \times \ \sqrt{\frac{63}{60} - \ \frac{-27}{60}}$$

$$= 10 \times \sqrt{2.717 - 0.2025}$$

$$= 10 \times 1.586 = 15.86$$

c. Co-efficient of variation (V) = $\frac{S}{X} \times \frac{00}{} = \frac{5.86}{70} \times \frac{00}{} = 22.66\%$

CLASS ASSESSMENT EXERCISE 3

1. The following table shows the distribution of masses of 100 blood donors.

Mass(kg)	60 – 62	63 – 65	66 – 68	69 – 71	72 – 74
No of donors	5	18	42	27	8

(a) Find the standard deviation.
(b) Using an assumed mean of 6.7kg; find the standard deviation.
(c) Estimate the co-efficient of variation.

2. Given a set of data as:

8	5	2	8	13	15	2	25	8	16
21	14	3	4	17	7	11	15	9	2

Find the standard deviation and the coefficient of variation.

4.0 CONCLUSION

This unit has successfully examined the various measures of dispersion along with the procedures involved in computing the estimates in both grouped and ungrouped data.

5.0 SUMMARY

Measures of dispersion are statistical estimates which show the degree of spread of observation in a given data from the arithmetic averages. Commonest among the measures are the range, mean deviation, standard deviation and variance. These estimates can be obtained for both grouped and ungrouped data following certain steps or procedures.

6.0 TUTOR MARKED ASSINMENT

1. Given the weight in (kg) of 20 kids as

5	6	8	10	7	8	5	4	6	8
7	9	6	5	8	4	6	2	5	4

Find the range, mean deviation, standard deviation, variance and the coefficient of variation.

2. The frequency distribution below show the maximum loads in kilonewtons (KN) supported by some cables provided by a certain company.

Maximum Load (kN)	Number of cables
4-9	3
10-15	5
16-21	12
22-27	17
28-33	14
34-39	6
40-45	3

Find: -

(a) the mean deviation
(b) the standard deviation
(c) the variance
(d) the standard deviation using the assumed mean of 24.5kN

7.0 REFERENCE/ FURTHER READING

Akanbi, S. O. and Chika, N. C. (2007). Mathematics for Senior Secondary School. Macmillan Publishing, Nigeria.

Channon, J. B. et al (2002). New General Mathematics for Senior Secondary Scholls 3, Longman Publishers, Nigeria.

Spiegel, M. R. and Stephens, L. T. (2009). Statistics. Schaums Outline Series (4th Edition), McGraw Hills Publishers, Delhi.

UNIT 4

MEASURES OF PARTITION

Table of Contents

1.0 INTRODUCTION

The measure of partitions refers to the statistical estimates obtained as a result of breaking data or other quantitative or qualitative information into groups, parts or divisions. When data is broken down into four equal parts or division, each part or division is called a *Quartile*. When data is broken down into ten equal parts or division, each part or division is called a *Decile*. In breaking a data into one-hundred equal parts, each portion or part is called a *Percentile*. Quartiles, deciles and percentiles are called the measures of partition and they are interrelated as follows: -

1st quartile = 2.5th Decile = 25th Percentile
2nd quartile = 5th Decile = 50th Percentile
3rd quartile = 7.5th Decile = 75th Percentile
4th quartile = 10th Decile = 100th Percentile

Similarly,

1st Decile = 10th Percentile
2nd Decile = 20th Percentile
3rd Decile = 30th Percentile
4th Decile = 40th Percentile
5th Decile = 50th Percentile = Median
6th Decile = 60th Percentile
7th Decile = 70th Percentile
8th Decile = 80th Percentile
9th Decile = 90th Percentile
10th Decile = 100th Percentile

2.0 OBJECTIVES

At the end of this unit, you must be able to: -

(i) Define the various measures of partition.
(ii) Calculate the various measures of partition for both grouped and ungrouped data.

3.0 MAIN CONTENT

3.1 Quartiles

A quartile is one part of a data when the data is divided into four equal parts. So, we have 1st quartile, 2nd quartile and 3rd quartile and 4th quartile. It should be noted that the difference between the 3rd quartile (Q_3) and 1st quartile (Q_1) is called the inter-quartile range.

$\therefore$ Inter-quartile range = $Q_3 - Q_1$,

while semi-interquartile range = $\frac{Q_3 - Q_1}{2}$

Where Q_3 = 3rd quartile or upper quartile; and
Q_1 = 1st quartile or lower quartile.

For Grouped data,

$$Q_x = \frac{X}{4}N + 0.5 \quad ah \text{ taRnt}$$

Where X = position of the quartile e.g. for 3rd quartile, we have:
$Q_3 = \frac{3}{4}N + 0.5 \quad ah \text{ taRnt}$
N = $\sum \mathbf{f}$ or total number of the observation
For grouped data; $Q_x = L_x + \frac{\frac{X}{4}N + m \quad c}{\mathbf{f}}$

Where; L_X = lower limit/boundary of the quartile class.
N = $\sum \mathbf{f}$ or number of the observation
X = position of the quartile
f = frequency of the quartile class
c = class interval
m = cumulative frequency just before the quartile class.

Worked Example

1. Given the set of observation as: -
7, 12, 18, 15, 20, 19, 16, 13, 23 and 17.
Find: -
(a) The lower quartile
(b) The semi-interquartile range

2. The ages of retired personnel in a local government area is given by the frequency distribution table as: -

Ages	**Frequency**
51 – 56	5
57 – 62	10
63 – 68	11
69 – 74	14
75 – 80	16
81 – 86	4

Obtain the interquartile range and the middle quartile (Q_2)

Solution

1. Arranging the data in ascending order of magnitude: -
7, 12, 13, 15, 16, 17, 18, 19, 20, 25

(a) $Q_1 = \frac{}{4}N + 0.5$ ah taRnt

$= \frac{}{4} \times 10 + 0.5$ ah taRnt

$= (2.5 + 0.5)$ah taRnt

$= 3\mathbf{f}$ taRnt $= 13$

Note: - 3 is not the Q_1 but the item in the 3rd position.

(b) Semi-interquartile range = $\frac{Q_3 - Q_1}{2}$

$Q_3 = \frac{3}{4}N + 0.5$ ah taRnt

$= (\frac{3}{4}(10) + 0.5)ah$ taRnt

$= (7.5 + 0.5)$ah taRnt

$= 8$ah taRnt $= 19$

$Q_1 = \frac{}{4}N + 0.5$ ah taRnt

$= \frac{}{4} \times 10 + 0.5$ ah taRnt

$= (2.5 + 0.5)$ah taRnt

$= 3\mathbf{f}$ taRnt $= 13$

∴ Semi interquartile range = $\frac{Q_3 - Q_1}{2} = \frac{9- 3}{2} = 3$

2.

Ages	Frequency	Cumm. Freq.	Class boundaries
51 – 56	5	5	50.5 – 56.5
57 – 62	10	15	56.5 – 62.5
63 – 68	11	26	62.5 – 68.5
69 – 74	14	40	68.5 – 74.5
75 – 80	16	56	74.5 – 80.5
81 – 86	4	60	80.5 – 86.5
	$\sum \mathbf{f} = 60$		

(i) Interquartile range = $Q_3 - Q_1$

Q3 = $\frac{3}{4}N + 0.5)ah\ aR\mathbf{f}nt$

$= \frac{3}{4} \times (60) + 0.5\ \ ah\ aR\mathbf{f}nt$

$= 45.5ah\ aR\mathbf{f}nt$. This implies that the Q_3 class is where the cumulative frequency of 45.5 falls, i.e. the class of 75 – 80.

$\therefore Q_3 = L_3 + \frac{\frac{3}{4}N + m\ \ c}{\mathbf{f}}$

$= 74.5 + \frac{\frac{3}{4}\times 60\ \ - 40\ \ 6}{6}$

$= 74.5 + \frac{45-40}{6}\ \ 6$

$= 74.5 + \frac{5}{6}\ \ 6$

$= 74.5 + \frac{30}{6}$

$= 74.5 + 1.875 = 76.375$

$Q_1 = \frac{-}{4}N + 0.5\ \ ah\ aR\mathbf{f}nt$

$= \frac{-}{4} \times (60) + 0.5\ \ ah\ aR\mathbf{f}nt$

$= (15 + 0.5)ah\ aR\mathbf{f}nt$

$= 45.5ah\ aR\mathbf{f}nt$. This implies that he Q_1 class is where the cumulative frequency of 15.5 lies (63 – 68).

$\therefore Q_1 = L\ \ + \frac{\frac{1}{4}N + m\ \ c}{\mathbf{f}}$

$= 62.5 + \frac{\frac{1}{4}\times 60\ \ -\ \ 5\ \ 6}{}$

$= 62.5 + \frac{5-\ \ 5}{}\ \ 6$

$= 62.5 + \frac{0}{6}\ \ 6$

$= 62.5 + (0)$

$= 62.5$

$\therefore$ Interquartile range = $Q_3 - Q_1$

= 76.375 – 62.5

= 13.875

(ii) Q_2 class = $\frac{-}{2} N + 0.5$ ah aRfnt

= $\frac{-}{2} \times (60) + 0.5$ ah aRfnt

= $(30 + 0.5)$ah aRfnt

= 30.5ah aRfnt. The corresponding frequency distribution is 69 – 74.

$$\therefore Q_2 = L_2 + \frac{\frac{2}{4}N + m \quad c}{f}$$

$$= 68.5 + \frac{\frac{2}{4} \times 60 \quad - 26 \quad 6}{4}$$

$$= 68.5 + \frac{30 - 26}{4} \quad 6$$

$$= 68.5 + \frac{4}{4} \quad 6$$

$$= 68.5 + .714$$

$$= 70.214$$

CLASS ASSESSMENT EXERCISE 1

1. Given the following as the ages of students in a class: -

   17 20 19 20 16 19 19 19 20 19 18 17

   17 19 20 20 20 21 20 25 23 23 22 30

   Find;

   (a) The upper quartile

   (b) Semi-interquartile range

2. The distribution of scores students in an examination is given as: -

Ages(yrs)	20–29	30–39	40–49	50–59	60–69	70–79	80–89	90–99
Freq.	9	10	13	19	11	8	6	4

Find (a) The lower quartile

(b) The interquartile range

(c) The semi-interquartile range

3.2 Deciles

A decile is one part when a distribution is broken down into ten equal parts or divisions. Hence, we have 1st decile, 2nd decile, 3rd decile, 4th decile ... 10th decile.

For an ungrouped data, the decile is given as: $D_x = \frac{X}{0} N + 0.5$ ah aRfnt

Where; X = position of the decile.

N = $\sum f$ or sum of all observation.

Note: - The item in the position is the decile not the position itself.

For grouped data, the decile is given as: $D_x = L_X + \frac{\frac{X}{10}N - m}{f}c$

Where; Lx = lower limit of the decile class
N = $\sum f$ or sum of all observation
c = class interval or class size
f = frequency of the decile class
X = position of the decile
m = cumulative frequency before the decile class

Worked Examples

1. Given a set of scores as: 17, 23, 13, 12, 16, 7, 19, 20, 18 and 15. Find the 8th decile.
2. Given the scores of students in an exam as:

Scores	**Frequency**
51 – 56	5
57 – 62	10
63 – 68	11
69 – 74	14
75 – 80	16
81 – 86	4

Find the 4th decile?

Solution

1. Rearranging the scores in ascending order

7, 12, 13, 15, 16, 17, 18, 19, 20, 23

8th decile = $\frac{8}{10} \times 10 + 0.5$ th item

$= (8 + 0.5)$th item $= 8.5$th item

$\therefore$8th decile = $\frac{8\text{th item} + 9\text{th item}}{2}$

$= \frac{9+20}{2} = 19.5$

2.

Scores	Freq.	Cum. Freq.	Class boundary
51 – 56	5	5	50.5 – 56.5
57 – 62	10	15	56.5 – 62.5
63 – 68	11	26	62.5 – 68.5
69 – 74	14	40	68.5 – 74.5
75 – 80	16	56	74.5 – 80.5
81 – 86	4	60	80.5 – 86.5

$$D_4 = \frac{4}{0} \times N + 0.5 \ \text{ah aRfnt}$$
$$= \frac{4}{0} \times 60 + 0.5 \ \text{ah aRfnt}$$
$$= (24 + 0.5)\text{ah aRfnt} = 24.5\text{ah aRfnt}$$

This falls in the class of 63 – 68 under the cumulative frequency distribution.

$\therefore$ 4th decile (D_4) = $L_4 + \frac{\frac{4}{10}N - m\ c}{f} = 62.5 + \frac{\frac{4}{10} \times 60 - 5\ 6}{} = 62.5 + \frac{24 - 5}{} 6 = 62.5 + \frac{9}{} 6 = 62.5 + 4.90 = 67.40$

CLASS ASSESSMENT EXERCISE 2

1. The data blow is the years of experience of selected senior staff of an organization: 21, 22, 23, 22, 21, 25, 28, 26, 25, 24, 23, 26, 28, 30, 27, 25, 24, 26, 21 and 27. Find;
   (a) The 7th decile (b) The 2nd decile
2. Given the distribution of students scores as:

Scores	1 – 20	21 – 40	41 – 60	61 – 80	81 – 100
Freq.	25	35	40	20	80

Find; (a) The 9th decile
(b) The 6th decile
(c) The 2nd decile

3.3 Percentile

A percentile is one part when a distribution is divided into one hundred (100) equal parts. Therefore, for a given data, we can obtain the 1st percentile, 2nd percentile, 3rd percentile ... 10th percentile, 20th percentile ... 100th percentile.

For an ungrouped data, percentile is given as:

$$P_X = \frac{X}{00} \times N + 0.5 \ \text{ah aRfnt}$$

Where; X = position of the percentile
N = $\sum f$ or sum of all observation.

It should be noted that, the formula above gives the position of the percentile and not the percentile itself.

For grouped data, percentile is given as: $P_x = L_X + \frac{\overline{100}N - m\ c}{f}$

Where; Lx = lower limit of the percentile class
N = $\sum f$ or sum of all observation
c = class interval or class size
f = frequency of the percentile class
X = position of the percentile
m = cumulative frequency before the percentile class

Percentile range could be obtained for two extreme ends of percentile such as:

$P_{90} - P_{10}$, $P_{80} - P_{20}$, $P_{70} - P_{30}$, etc.
10 – 90 percentile range = $P_{90} - P_{10}$
20 – 80 percentile range = $P_{80} - P_{20}$, etc.

Worked Example

1. The ages of selected ten members of the same family is given as: 3, 5, 2, 6, 5, 9, 5, 2, 8, 6.
   (a) Prepare a frequency distribution table for the distribution table for the distribution and find the 70th percentile.
   (b) Find 10 – 90 percentile range?

Solution

(a)

Ages	**Tally**	**Freq. (f)**	**Cum. Freq.**
2	ll	2	2
3	l	1	3
5	lll	3	6
6	ll	2	8
8	l	1	9
9	l	1	10
		$\sum f$ = 10	

70th percentile = $\frac{70}{00} \times N + 0.5$ ah taRnt

$= \frac{70}{00} \times 10 + 0.5$ ah taRnt

$= (7 + 0.5)$ah taRnt

$= 7.5ah\ taRnt = \frac{7ttt\ item + 8ttt\ item}{2}$

$= \frac{6+6}{2} = \frac{2}{2} = 6$

(b) $P_{90} = \frac{90}{00} \times N + 0.5\ \ ah\ taRnt$

$= \frac{90}{00} \times 10 + 0.5\ \ ah\ taRnt$

$= (9 + 0.5)ah\ taRnt$

$= 9.5ah\ taRnt = \frac{9ttt\ item + \ 0ttt\ item}{2}$

$= \frac{8+9}{2} = \frac{7}{2} = 8.5$

$P_{10} = \frac{0}{00} \times N + 0.5\ \ ah\ taRnt$

$= \frac{0}{00} \times 10 + 0.5\ \ ah\ taRnt$

$= (1 + 0.5)ah\ taRnt$

$= 1.5ah\ taRnt = \frac{t\ item + 2nd\ item}{2}$

$= \frac{2+2}{2} = \frac{4}{2} = 2$

10 – 90 percentile range = $P_{90} - P_{10}$

= 6 – 2 = 4

2. Given the age distribution of 60 retirees as:

Ages(yrs)	51 – 56	57 – 62	63 – 68	69 – 74	75 – 80	81 - 86
Freq.	5	10	11	14	16	4

Find; (a) The 80th percentile.
(b) The 20th percentile.
(c) The 20 – 80 percentile range.

Solution

Scores	**Freq.**	**Cum. Freq.**	**Class boundary**
51 – 56	5	5	50.5 – 56.5
57 – 62	10	15	56.5 – 62.5
63 – 68	11	26	62.5 – 68.5
69 – 74	14	40	68.5 – 74.5
75 – 80	16	56	74.5 – 80.5
81 – 86	4	60	80.5 – 86.5
	$\sum \mathbf{f} = 60$		

(a) P_{80} class = $\frac{80}{00} \times \mathrm{N} + 0.5$ ah aR**f**nt

= $\frac{80}{00} \times 60 + 0.5$ ah aR**f**nt

= 48.5ah aR**f**nt. The class group/age group corresponding to this is 75 – 80 class.

$\therefore P_{80} = L_{80} + \frac{\left(\frac{80}{100}N - m\right)c}{f}$

$= 74.5 + \frac{\left(\frac{80}{100}\times 60 - 40\right)6}{6}$

$= 74.5 + \frac{48-40}{6}\, 6$

$= 74.5 + \frac{8}{6}\, 6$

$= 74.5 + 6 = 77.5$

(b) P_{20} class = $\frac{20}{00} \times N + 0.5$ ah aRfnt

$= \frac{20}{00} \times 60 + 0.5$ ah aRfnt

$= 12.5$ah aRfnt. The class group/age group corresponding to this is 51 – 62 class.

$\therefore P_{20} = L_{20} + \frac{\left(\frac{20}{100}N - m\right)c}{f}$

$= 56.5 + \frac{\left(\frac{20}{100}\times 60 - 5\right)6}{0}$

$= 56.5 + \frac{20-5}{0}\, 6$

$= 56.5 + \frac{7}{0}\, 6 = 56.5 + \frac{42}{0}$

$= 56.5 + 4.2 = 60.7$

20 – 80 percentile range = $P_{80} - P_{20}$

= 77.5 – 60.7

= 16.8

CLASS ASSESSMENT EXERCISE

1. Given the scores of students in a test as:
   6, 8, 5, 6, 7, 8, 6, 4, 3, 2, 1, 4, 5, 6, 8, 7, 8, 6, 7, 6, 5, 6, 8, 7, 6, 5, 4, 6, 9, and 10. Find P_{90}, P_{10}, and 10 – 90 percentile range?
2. Given the heights of 40 athletes as:

Heights (cm)	150 – 159	160 – 169	170 –179	180 – 189	190 – 199
Freq.	25	45	30	40	60

Find (a) The 30th percentile.
(b) 10 – 90 percentile range.

4.0 CONCLUSION

The unit examined the meaning, concept, scope and the computation of the various measures of partition.

5.0 SUMMARY

Measures of partition of are statistical estimates that describes the membership of a group or division when a distribution is divided into a given equal parts. Deciles, quartile and percentiles are the commonly used measure of partition. Deciles are the estimate obtained when data is broken down into ten equal parts, quartiles are the ones obtained when a distribution is divided into four equal parts while percentiles are obtained when data is broken down into 100 equal parts. Deciles, quartile and percentiles can be obtained for both grouped and ungrouped data.

6.0 TUTOR MARKED ASSIGNMENT

1) The data below gives the scores of students in a quiz:

3	3	4	3	8	4	12	11	10	3
4	5	8	8	11	12	8	5	7	12
10	12	12	11	10	8	7	7	7	7

Find: - (a) The semi interquartile range
(b) The 9 decile
(c) The 40 percentile
(d) The 10 – 90 percentile range.

2) Given the monthly rent in naira of 200 households in thousand naira as: -

Rent(N'000)	**No of households**
8 – 12	12
13 – 17	26
18 – 22	45
23 – 27	60
28 – 32	37
33 – 37	13
38 – 42	5
43 – 47	2

Find: - (i) Interquartile range
(ii) Middle quartile
(iii) 7th decile
(iv) 20 – 80 percentile range

7.0 REFERENCES/FURTHER READINGS

Channon, J. B. et al (2012). New General Mathematics for Senior Secondary Schools, Book 1 – 3. Longman Publishers, Nigeria.

Spiegel, M. R. and Stephens, L. J. (009). Statistics, Schaums Outline Series (4th Edition). McGraw Hills Publishers, Delhi.

Module 4: Moments, Skewness and Kurtosis

Unit 1: Moments

Unit 2: Skewness

Unit 3: Kurtosis

UNIT 1: MOMENTS

Table of Contents

1.0 INTRODUCTION

In the last module, we learnt about how to find the degree of variation which is mainly what the concept of dispersion is all about. In this module, we shall learn about the direction of this variation which is known as Moments.

2.0 OBJECTIVES

At the end of the unit, you should be able to: -

i. Define the concept of Moment;
ii. Compute first, second, third and fourth moments of ungrouped data;
iii. Compute first, second, third and fourth moments of grouped data;
iv. Apply Charlier's check and Sheppard's correction in computing Moments; and
v. Define Moments in Dimensionless form and establish relationship between Moments.

3.0 MAIN CONTENT

3.1 General Overview

Moments is a general class of measures used in measuring the central tendency and dispersion of variable x. The $\mathbf{f}^{ttt}$ moment of a distribution with variables $X_1, X_2, X_3 \ldots X_n$ is defined as the arithmetic means of powers of r. This is denoted by

$$\bar{X} = \frac{x_1^r + x_2^r + x_3^r + \cdots x_n^r}{N} = \frac{\sum X^r}{N} \text{, for ungrouped data or}$$

$$\frac{\sum f x^r}{N} = \frac{\sum f x^r}{\sum f} \text{, for grouped data.}$$

The most general measure of this type is $M^r = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)$ where the raw moment r = 0, 1, 2 … k, which is called the r^{th} moment of x about the origin 'a'.

If the given values are classified into a frequency table or grouped data, it takes the form $M^r = \frac{1}{n}\sum_{i=1}^{k}(x_i - a)$, where x_i is the classmark or the midpoint of the i^{th} class and f_i is the frequency.

When the origin of a moment is taken as the arithmetic mean of the variable, it is called a ***Central Moment***. Thus, the r^{th} central moment denoted by M^r, is

$$M^r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r\text{: ungrouped data}$$

$$M^r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})\, f_i\text{: grouped data}$$

The mean of a variable is its first moment about zero, while the variance is the second central moment. Hence, we say that moment deals with the direction of variation.

Sometimes, the r^{th} central moment denoted by M^r, is given as:

$$M^r = \frac{\sum(x-\bar{x})^r}{N} \text{ or } \overline{(x-\bar{x})^r}\text{: ungrouped data}$$

and $M^r = \frac{\sum_{j=1}^{n}(x_j - A)^r}{N} = \frac{\sum(X-A)^r}{N} = \frac{\sum d^r}{N} = \overline{(X-\bar{X})^r}$ for grouped data or ungrouped data involving assumed mean where d = X – A are the derivations from X from A. if A = 0, M^r reduces to $\frac{\sum X^r}{N}$, where is often called ***moment about zero***.

Specifically, for grouped data, if X_1, X_2 …X_k occur with frequency f_1, f_2, f_3 …f_k, respectively,

$$\bar{X} = \frac{f_1X_1^r + f_2X_2^r + \cdots + f_kX_k^r}{N} = \frac{\sum_{j=1}^{k} f_j X_j^r}{N} = \frac{\sum fX^r}{N}$$

Hence,

$$M^r = \frac{\sum_{j=1}^{k} f_j (X_j - \overline{X})^r}{N} = \frac{\sum f(X - \overline{X})^r}{N} = \overline{(X - \overline{X})^r}$$

and $M^r = \sum_{j=1}^{k} f_j (X_j - A)^r = \frac{\sum f(X-A)^r}{N} = \overline{(X-A)^r}$

Where A is the assumed mean and X – A are the deviations of X from A.

Note: $N = \sum f$

CLASS ASSESSMENT EXERCISE 1

1) Define the term 'Moment'. How is it different from Measures of Dispersion?
2) Define each of the following:
   (a) Origin of Moment
   (b) Moment about Zero
3) State the alternative mathematical expression for
   (a) Moment of Ungrouped Data
   (b) Moment of Grouped Data
   (c) Moment of Ungrouped Data using assumed mean (A)
   (d) Moment of Grouped Data using assumed mean (A)

3.2 Moment of Ungrouped Data

Worked Examples

1. Find (a) the first, (b) the second, (c) the third and (d) fourth moments of the set of numbers: 2, 3, 4, 5, 6.

Solution

(a) The first moment = $\frac{\sum X}{N} = \bar{X}$

$\bar{X} = \frac{2+3+4+5+6}{5} = \frac{20}{5} = 4$

(b) The second moment is $\frac{\sum X^2}{N} = \frac{2^2+3^2+4^2+5^2+6^2}{5} = \frac{90}{5} = 18$

(c) The third moment is $\frac{\sum X^3}{N} = \frac{2^3+3^3+4^3+5^3+6^3}{5} = \frac{8+27+64+\quad 25+2\quad 6}{5} = \frac{440}{5} = 88$

(d) The fourth moment is $\frac{\sum X^4}{N} = \frac{2^4+3^4+4^4+5^4+6^4}{5} = \frac{6+8\quad +256+625+\quad 296}{5}$

$= \frac{2274}{5} = 454.8$

2. Find the first three moments about the 7 of 2, 3, 5, 6 and 9.

Solution

(a) The first moment about 7 is

$\frac{(2-7)+(3-7)+(5-7)+(6-7)+(9-7)}{5} = \frac{-5+(-4)+(-2)+(-\quad)+2}{5} = \frac{-\quad 0}{2} = -2$

(b) The second moment about 7 is

$$\frac{(2-7)^2+(3-7)^2+(5-7)^2+(6-7)^2+(9-7)^2}{5} = \frac{25+\ \ 6+4+\ \ +4}{5} = \frac{50}{5} = 10$$

The third moment about 7 is

$$\frac{(2-7)^3+(3-7)^3+(5-7)^3+(6-7)^3+(9-7)^3}{5} = \frac{-\ \ 25-64-8-\ \ +8}{5} = \frac{-\ \ 90}{5} = -38$$

3. Find the first four moments about the mean of 2, 3, 7, 8 and 10.

Solution

The mean $= \frac{2+3+7+8+\ \ 0}{5} = \frac{30}{5} = 6$

(a) The first moment about the mean

$$= \frac{(2-6)+(3-6)+(7-6)+(8-6)+(\ \ 0-6)}{5} = \frac{-4+(-3)+\ \ +2+4}{5} = \frac{0}{5} = 0$$

Thus the first moment about the mean is zero.

(b) The second moment about the mean is $\frac{\Sigma(X-\bar{X})^2}{N}$

$= \frac{(2-6)^2+(3-6)^2+(7-6)^2+(8-6)\ \ \ +(\ \ 0-6)^2}{5} = \frac{\ 6+9+\ \ +4+\ \ 6}{5} = \frac{46}{5} = 9.2$. This is the Variance of the distribution.

(c) The third moment about the mean is $\frac{\Sigma(X-\bar{X})^3}{N}$

$$= \frac{(2-6)^3+(3-6)^3+(7-6)^3+(8-6)^3+(\ \ 0-6)^3}{5} = \frac{-64-27+\ \ +8+64}{5} = \frac{-\ \ 8}{5} = -3.6$$

(d) The fourth moment about the mean = $\frac{\Sigma(X-\bar{X})^4}{N}$

$$= \frac{(2-6)^4+(3-6)^4+(7-6)^4+(8-6)^4+(\ \ 0-6)^4}{5} = \frac{256+8\ \ +\ \ +\ \ 6+256}{5} = \frac{6\ \ 0}{5} = 122$$

CLASS ASSESSMENT EXERCISE 2

1. Find the (a) first, (b) second, (c) third and (d) fourth moments of the set: 2, 3, 7, 8 and 10.
2. Find the first moment about the mean of 1, 2, 4, 6 and 7.
3. Given the set of 2, 3, 7, 8 and 10; find:
   (a) M_1 (1^{st} moment about the mean)
   (b) M_2 (2^{nd} moment about the mean)
   (c) M_3 (3^{rd} moment about the mean)
   (d) M_4 (4^{th} moment about the mean)
4. For question (3) above, find the first, second, third and fourth moment about the origin. (M_1,M_2, M_3,M_4)

3.3 Moment of Grouped Data

Example 1

Given the table below:-

Group	10-12	13-15	16-18	19-21
Frequency	3	1	4	2

Required: Find the first four moments about the mean.

Solution

Class Group	Midmark (x)	f	fx	$(x-\bar{x})$	$f(x-\bar{x})$	$f(x-\bar{x})^2$	$f(x-\bar{x})^3$	$f(x-\bar{x})^4$
10-12	11	3	33	-4.5	-13.5	60.75	-273.375	1230.1875
13-15	14	1	14	-1.5	-1.5	2.25	-3.375	5.0625
16-18	17	4	68	1.5	6.0	9.0	13.5	20.25
19-21	20	2	40	4.5	9.0	40.5	182.25	820.125
Total		10	155		0	112.5	-81	2075.625

Mean $(\bar{X}) = \frac{\sum fX}{\sum f} = \frac{55}{0} = 15.5$

(a) The first moment about the mean $= \frac{\sum f(X-\bar{X})}{\sum f} = \frac{0}{0} = 0$

(b) The second moment about the mean $= \frac{\sum f(X-\bar{X})^2}{\sum f} = \frac{2.5}{0} = 11.25$

(c) The third moment about the mean $= \frac{\sum f(X-\bar{X})^3}{\sum f} = \frac{-8}{0} = -8.1$

(d) The third moment about the mean $= \frac{\sum f(X-\bar{X})^4}{\sum f} = \frac{2075.625}{0} = 207.5625$

Example 2

Given the following data below:-

Height (inches)	**Frequency**
60-62	5
63-65	18
66-68	42
69-71	27
72-74	8
	$\sum f = 100$

Compute

(a) M^I, M^I_2, M^I_3, and M^I_4 (first four moment about the origin)

(b) M_1,M_2, M_3, and M_4 (first four moment about the mean)

Hint: use assumed mean of 54.

Solution

Class Group	Midpoint (X)	u = X-A	f	fu	fu^2	fu^3	fu^4
60–62	61	-2	5	-10	20	-40	80
63–65	64	-1	18	-18	18	-18	18
66–68	67	0	42	0	0	0	0
69–71	70	1	27	27	27	27	27
72–74	73	2	8	16	32	64	128
			$N=\sum f=10$	$\sum fq=15$	$\sum fu^2=97$	$\sum fu^3=33$	$\sum fu^4=25$

(a) (i) $M^I = c\frac{\sum fu}{N} = 3\ \frac{5}{00} = 0.45$

(ii) $M_2^I = c^2 \frac{\sum fu^2}{N} = (3^2)\frac{97}{00} = 8.73$

(iii) $M_3^I = c^3 \frac{\sum fu^3}{N} = (3^3)\frac{33}{00} = 8.91$

(iv) $M_4^I = c^4 \frac{\sum fu^4}{N} = (3^4)\frac{253}{00} = 204.93$

(b)(i) $M_1 = 0$

(ii) $M_2 = M_2^I - M^{I\,2} = 8.73 - (0.45)^2 = 8.5272$

(iii) $M_3 = M_3^I - 3M^I M_2^I + M^{I\,3}$

$= 8.91 - 3(0.45)(8.73) + 2(0.45)^3 = 2.6932$

(iii) $M_4 = M_4^I - 4M^I M_3^I + 6M^{I\,2} M_2^I - 3M^{I\,4}$

$= 204.93 - 4(0.45)(8.91) + 6(0.45)^2(8.73) - 3(0.45)^4$

$= 199.3759$

Note: You can use the formula for moment about the mean as we do in example 1 (unit 3.3) instead of using the relationship identified in (b) above i.e.

$\frac{\sum((X-\bar{X})^r)}{\sum f}$, where x = midpoint

It should be noted that for grouped data involving assumed mean,

$\bar{X} = A + \frac{\sum fd}{\sum f}$, where A = assumed mean and d = (X-A).

CLASS ASSESSMENT 3

Given the table below of a discrete frequency distribution, find the

(a) first three moments about the mean
(b) first three moments about the origin
(c) first three moment about the mean given that assumed mean is 16

Group	4-7	8-11	12-15	16-19
Frequency	2	2	2	2

3.4 Charlier's Check and Sheppard's Correction

Charlier's check and Sheppard's correction are used to guide against computational error. In doing this (either for grouped or ungrouped data) additional column could be added to the table. This results in Charlier's check.

Charlier's check in computing moments by the coding method uses the identities:

$$\sum \mathbf{f}(q+1) = \sum \mathbf{f}q + N$$
$$\sum \mathbf{f}(q+1)^2 = \sum \mathbf{f}q^2 + 2\sum \mathbf{f}q + N$$
$$\sum \mathbf{f}(q+1)^3 = \sum \mathbf{f}q^3 + 3\sum \mathbf{f}q^2 + 3\sum \mathbf{f}q + N$$
$$\sum \mathbf{f}(q+1)^4 = \sum \mathbf{f}q^4 + 4\sum \mathbf{f}q^3 + 6\sum \mathbf{f}q^2 + 4\sum \mathbf{f}q + N$$

Note: - The coefficient for the expansion follows the Paschal triangle.

1 1

1 2 1

1 3 3 1

1 4 6 4 1, and so on.

The Sheppard's corrections for moments are as follows:

$$\text{Corrected } M_2' = M_2' + \frac{}{2}C^2$$

$$\text{Corrected } M_4' = M_4' + \frac{}{2}C^2 M^2 + \frac{7}{240}C^3$$

Example 1

Given the table below:-

X	70	74	78	82	86	90	94	98	102	106	110	114	118	122	126
F	4	9	16	28	45	66	85	72	54	38	27	18	11	5	2

Using an assumed mean of 94 (i.e. A = 94), check whether the Charlier's check hold or not.

Solution

X	u	F	f(u+1)	$f(U+1)^2$	$f(U+1)^3$	$f(U+1)^4$
70	-6	4	-5	100	-500	2500
74	-5	9	-4	144	-576	2304
78	-4	16	-3	144	-432	1296
82	-3	28	-2	112	-224	448
86	-2	45	-1	45	-45	45
90	-1	66	0	0	0	0
94	0	85	1	85	85	85
98	1	72	2	288	576	1152
102	2	54	3	486	1458	4374
106	3	38	4	608	2432	9728
110	4	27	5	675	3375	16875
114	5	18	6	648	3888	23328
118	6	11	7	539	3773	26411
122	7	55	8	320	2560	20480
126	8	2	9	162	1458	13122
		$N = \sum \mathbf{f}$ =480	$\sum f(u+1)$ = 716	$\sum f(u+1)^2$ =4356	$\sum f(u+1)^3$ =17828	$\sum f(u+1)^4$ =122148

$\sum f(u+1)$ =716

$\sum \mathbf{f}q$ + N = 236 + 480 = 716

: - $\sum f(u+1) = \sum \mathbf{f}q$ + N = 716

$\sum f(u+1)^2$ = 4356

$\sum \mathbf{f}q^2 + 2\sum \mathbf{f}q$ + N = 3403 +2(236) + 480 = 4356

$\sum \mathbf{f}(q+1)^3$ = 17,828

: - $\sum f(u+1)^2 = \sum \mathbf{f}q^2 + 2\sum \mathbf{f}q$ + N

$\sum \mathbf{f}q^3 + 3\sum \mathbf{f}q^2 + 3\sum \mathbf{f}q$ + N = 6428 + 3(3404) + 3(236) + 480 = 1782

: - $\sum \mathbf{f}(q+1)^3 = \sum \mathbf{f}q^3 + 3\sum \mathbf{f}q^2 + 3\sum \mathbf{f}q$ + N

$\sum \mathbf{f}(q+1)^4$ = 122,148

$\sum \mathbf{f}q^4 + 4\sum \mathbf{f}q^3 + 6\sum \mathbf{f}q^2 + 4\sum \mathbf{f}q$ + N

= 74588 + 4(6428) + 6(3403) + 4(236) + 480 = 122,148

: - $\sum \mathbf{f}(q+1)^4 = \sum \mathbf{f}q^4 + 4\sum \mathbf{f}q^3 + 6\sum \mathbf{f}q^2 + 4\sum \mathbf{f}q$ + N

Hence, the Charlier's check confirms that no computational error is found in the moments about the mean for the data.

Example 2

Apply the Sheppard's correction to determine the moments about the mean for the data treated in Example 1 above.

Given $M^2 = 8.5272$, $C_1 = 3$, $M^4 = 199.3759$

Corrected $M_2 = M^2 - \frac{C^2}{2} = 8.5272 - \frac{3^2}{2} = 7.78$

Corrected $M_4 = M_4 - \frac{}{2}C^2M_2 + \frac{7}{240}C^2$

$= 199.3759 - \frac{}{2}3^2(8.2575) + \frac{7}{240}$

$= 163.3646$

Therefore, M_1 and M_2 need no correction.

CLASS ASSESSMENT EXERCISE 4

Given the layout of a data in a table below:-

Class boundaries	Frequency
6.45-7.25	4
7.25-8.05	5
8.05-8.85	17
8.85-9.65	34
9.65-10.45	36
10.45-11.25	29
11.25-12.05	16
12.05-12.85	6
12.85-13.65	3

Using the class of 8.85-9.65 as A, use the Charlier's check to check for the accuracy of the computation on the table.

3.5 Basic Relation in Moments

The following relations exist between moments about the mean M_r and moments about an arbitrary origin M'_r.

$$M_2 = M'_2 - (M')^2$$
$$M_3 = M'_3 - 3M'M'_2 + 2(M')^3$$
$$M^4 = M'_4 - 4M'M'_3 + 6(M')^2M'_2 - 3(M')^4\text{, etc.}$$

Note: $M' = \bar{X} - A$

To avoid particular unit, we can define the dimensionless moments about the mean as:

$$l_r = \frac{M_r}{s^r} = \frac{M_r}{\sqrt{(M_2)^r}} = \frac{M_r}{\sqrt{M_2^r}}$$

where $s = \sqrt{M_2}$ is the standard deviation since $M_1 = 0$ and $M_2 = s^2$, we have $l_1 = 0$ and $l_2 = 1$

4.0 CONCLUSION

The unit presented a general preview of what moments of a set of data implies. It exposes the students to the computational procedures involved in obtaining the moments of both ungrouped and grouped data. Charlier's check and Sheppard's correction were also introduced to expose the students to the elementary approaches to checking computational errors in moment.

5.0 SUMMARY

Moment refers to the direction of variation in a set of data. Moments can be obtained about the origin as well as about the mean. The first moment about the origin is the mean while the first moment about the mean is zero. The second moment about the mean is the variance. In asymmetrical distribution, all odd moments about the mean is zero.

The r^{th} moment about origin is given as $\frac{\sum X^r}{N}$ (ungrouped data) and $\frac{\sum fX^r}{\sum f}$ (grouped data)

While the r^{th} moment about origin A is $\frac{\sum (X-A)^r}{N}$ for a set of ungrouped data and $\frac{\sum f(X-A)^r}{\sum f}$ for a frequency distribution (grouped data). The r^{th} moment about the mean is $\frac{\sum (X-\bar{X})^r}{N}$ or $\frac{\sum f(X-\bar{X})^r}{\sum f}$.

6.0 TUTOR MARKED ASSIGNMENT

1. The number of bags of rice sold at Daleko market by 50 traders is shown below:

Bags of Rice	1–5	6–10	11–15	16–20	21–25	26–30
No of Traders (freq)	7	10	16	9	5	3

(a) Find the first four moments about the origin, 0
(b) Find the first four moments about the origin 18
(c) Find the first four moments about the mean

2. Given the information below: -

X	f
10	4
11	6
12	7
13	3
14	2
15	1
16	1
17	6
18	10

(a) Find M ; M_2; M_3; M_4;

(b) Find M ; M_2; M_3; M_4;

(c) Use Charlier's check and the Sheppard's correction to establish whether there is error in the computation of the moment or not.

7.0 REFERENCES/FURTHER READINGS

Loto, M. A., Ademola, A. A. and Toluwase, J. A. (2008). Statistics Made Essay. Concept Publication Limited, lagos, Nigeria.

Spiegel, M. R. and Stephens, L. T. (2007). Statistics, Fourth Edittion. Schaum's Outline. McGraw Hill, Singapore.

UNIT 2: SKEWNESS

Table of Contents

1.0 INTRODUCTION

An important feature dealing with the difference in sample distributions is with respect to skewness. When histogram ir graphs of frequency curves or frequency polygons are plotted, the shape can generally be of two types. They may be symmetrical (bell-shaped) in which case the values are ranged systematically around the central maximum or they may be symmetrical. If they are not symmetrical, then, the concept of skewness comes into play.

2.0 OBJECTIVES

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

(i) Define the term "skewness";

(ii) Identify the various features of skewness; and

(iii) Describe, compute and interpret various measures of skewness

3.0 MAIN CONTENT

3.1 General Overview

Skewness is the degree of asymmetry or departure from symmetry of a distribution. If the frequency curve (smoothed frequency polygon) of a distribution has a longer tail to the right of the central maximum than to the right, the distribution is said to be slewed to the right, or have positive skewness. If the reverse is the case, it is said to be skewed to the left, or negatively skewed.

Skewness is thus, defined as a measure of departure of a distribution from symmetry. A distribution that is not symmetrical is asymmetrical or skew distribution. Skewness can also be defined as a measure of asymmetric of a distribution. For positive skewness, the standard deviation runs up to high scores. This happens in a salary structure where people with high salary are few; or in a difficult test where there are few high marks. However, negative skewness happens in the case of an easy test where many students have high scores.

For a symmetrical distribution, the mean, median and mode coincide. But in asymmetrical distribution, they do not. For positively skewed distribution, the mean is greater than the median and the median is greater than the mode i.e. mean > median > mode. However, for a negatively skewed distribution, the mean is less than median and the median is less than the mode.

Mean < median < mode.

Positive skewness occurs when the mean is increased by some abnormally high values while negative skewness occurs when the mean is reduced by abnormally low values.

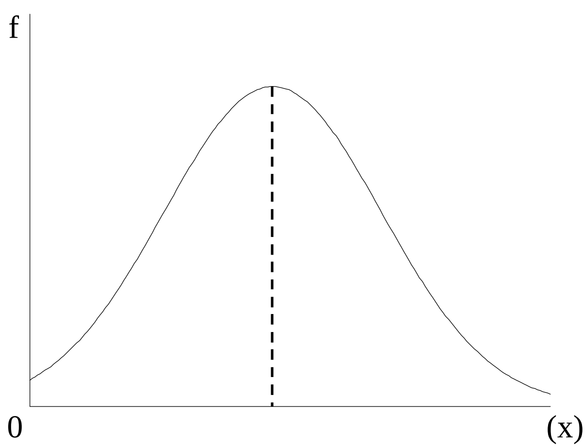
CLASS ASSESSMENT EXERCISE 1

1. Define the term "skewness".
2. Distinguish clearly between symmetrical distribution and asymmetrical distribution.

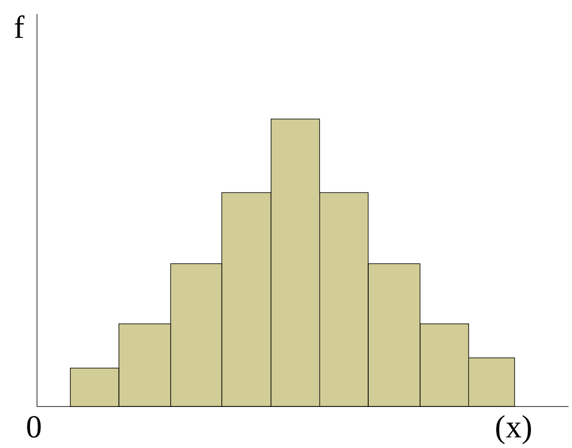
3.2 Symmetrical Distribution and Skewness

An important point which is to be noted in the connection of positive and negative skewness is that all odd-order central moments are zero for symmetrical distribution, positive for a positively skewed distribution and negative for a negatively skewed distribution. The diagram below illustrates symmetric distribution positively skewed and negatively skewed distribution.

Frequency Curve

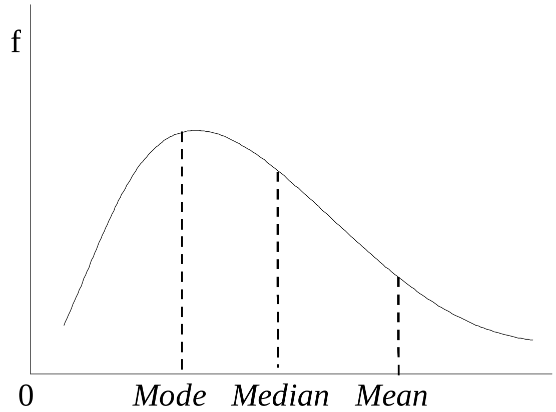


Median = Mean = Mode

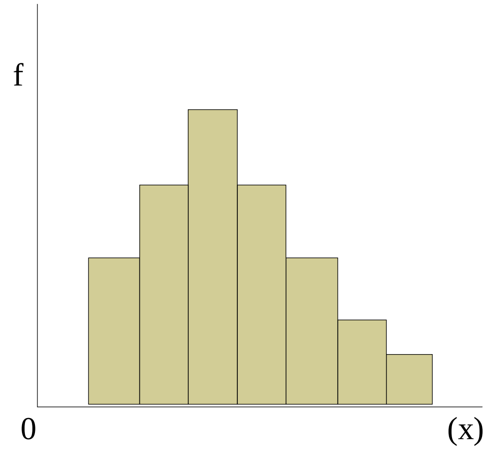


Histogram

In the figure above, all observations are equidistant from the mode. Hence, a symmetrical distribution has a zero skewness.

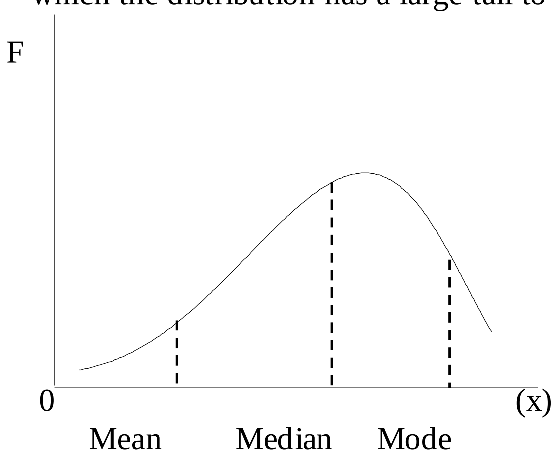


Frequency Curve

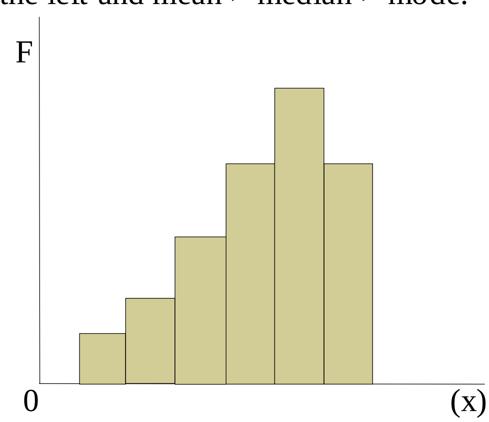


Histogram

The immediate frequency curve and histogram illustrate positive skewness in which the distribution has a large tail to the left and mean > median > mode.



Frequency Curve



Histogram

In the last diagram above, the frequency curve and the histogram illustrate positively skewed distribution. The longer tail is found at the right hand side of the frequency curve and mean < median < mode.

The skewness of a distribution could be described with respect to its direction and degree and this is represented by a computed co-efficient of skewness. By direction, we mean whether positively or negatively skewed. By degree, we mean whether the numerical value which is usually less than one. A normal (symmetrical) distribution has a co-efficient of skewness of zero; a co-efficient of +0.8 is positive and high while a co-efficient of -0.3 is negative and low.

CLASS ASSESSMENT EXERCISE 2

1. With the aid of graph (frequency distribution curve and histogram), distinguish clearly between symmetrical distribution, positively skewed distribution and negatively skewed distribution.
2. With appropriate examples and illustration, distinguish clearly between the degree and direction of skewness.

3.3 Measurement and Interpretation of Skewness

For skewed distributions, the mean tends to lie on the same side of the mode. Thus, a measure of the asymmetry is supplied by the difference between the mean and the mode: (mean – mode). This can be made dimensionless if we divide it by a measure of dispersion, such as the standard deviation.

The following are some of the dimensional approaches or alternative measures of skewness: -

(i) Skewness = $\frac{\text{mean}-\text{mode}}{\text{tandard deviation}} = \frac{\text{x}-\text{mode}}{}$

Recall that mean – mode = 3(mean – median)

(ii) Skewness = $\frac{3(\text{mean}-\text{median})}{\text{tandard deviation}}$

Equation in (ii) above is called the Pearson's first co-efficient of skewness while equation in (ii) is the Pearson's second co-efficient of skewness.

(iii) Bowley's quartile Co-efficient of skewness is given as:

$\frac{(Q_3-Q_2)-(Q_2-Q_1)}{Q_3-Q_1}$, opening the brackets of the numerator, we have:

$\frac{Q_3-Q_2-Q_2+Q_1}{Q_3-Q_1} = \frac{Q_3-2Q_2+Q_1}{Q_3-Q_1}$

(iv) 10 – 90 percentile Co-efficient of skewness, this is given as:

$\frac{(P_{90}-P_{50})-(P_{50}-P_{10})}{P_{90}-P_{10}}$, opening the brackets of the numerator, we have:

$\frac{P_{90}-P_{50}-P_{50}+P_{10}}{P_{90}-P_{10}} = \frac{P_{90}-2P_{50}+P_{10}}{P_{90}-P_{10}}$

(v) An important measure of skewness uses the third moment about the mean expressed in dimensional form and is given by

Moment co-efficient of skewness = $\mathbf{l}_3 = \frac{M_3}{S^3} = \frac{M_3}{(\mathbf{j}\overline{M_2})^3} = \frac{M_3}{\overline{M_2^3}}$

Where; S = standard deviation

S^2 = variance

$\mathbf{j}\overline{M_2}$ = variance

M_2 = second moment

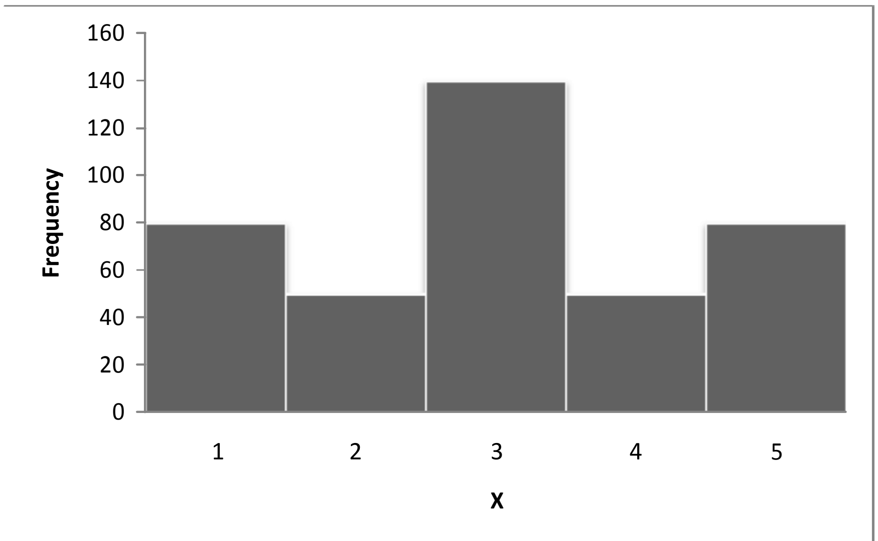
Another measure of skewness is sometimes given by $b_2 = \mathbf{l}_3^2$. For perfectly symmetrical curves, such as the normal curve, $\mathbf{l}_3$ and b_3 are zero.

Worked Examples

1. Comment on the skewness of the distribution of the table below:

X	1	2	3	4	5
F	80	50	140	50	80

Solution



This distribution is symmetrical since the observations are equidistant from the central maximum. Therefore, the skewness is zero.

2. The second moment about the mean of two distribution are 180 and 300 and the third moments bout the mean of two distribution are -91 and -150 respectively. Which of the two distributions is more skewness to the left?

Solution

1st distribution, $M_2 = 180, M^3 = -91$

∴ Moment co-efficient of skewness = $\frac{M_3}{(\sqrt{M_2})^3} = \frac{-9}{(\sqrt{80})^3} = -0.038$

2nd distribution, $M_2 = 300, M^3 = -150$

∴ Moment co-efficient of skewness = $\frac{M_3}{(\sqrt{M_2})^3} = \frac{-50}{(\sqrt{300})^3} = -0.029$

This implies that the first distribution (with moment co-efficient of skewness = -0.038) is more skewed to the left compared to the other distribution (-0.038 is more skewed to the left than -0.029).

3. Given the mean of a distribution as 60 and the mode as 50. If the standard deviation is 25. Find the co-efficient of skewness?

Solution

$$S_k = \frac{\text{mean} - \text{mode}}{\text{tandard deviation}} = \frac{60-50}{25} = \frac{0}{25} = 0.4$$

4. If the lower and upper quartiles of a distribution whose median is 1.56 are 0.48 and 2.37 respectively. Compute a co-efficient of skewness for the distribution.

Solution

The appropriate formula to use is the Bowley's co-efficient of skewness given by:

$$\frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} = \frac{2.37 - 2\times\ .56 + 0.48}{2.37 - 0.48} = \frac{2.37 - 3.\ 2 + 0.48}{.89} = -0.14$$

5. From a set of observation, mean = \$279.76, median = \$279.06, mode = \$277.50 and standard deviation = s = \$15.60. Find and interpret;
   a. First co-efficient of skewness
   b. Second co-efficient of skewness
   c. Quartile co-efficient of skewness
   d. Percentile co-efficient of skewness

Solution

(a) First co-efficient of skewness = $\frac{\text{mean} - \text{mode}}{\text{tandard deviation}} = \frac{\$279.96 - \$277.50}{\$\ 5.60}$

$= 0.1448 \text{ of } 0.15$

(b) Second co-efficient of skewness = $\frac{3(\text{mean} - \text{median})}{\text{tandard deviation}} = \frac{3(\$279.76 - \$27906)}{\$\ 5.60}$

$= 0.1346 \text{ of } 0.14$

Since the co-efficients are positive, the distribution is skewed positively (i.e. to the right).

(c) Quartile co-efficient of skewness = $\frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} = \frac{\$290.75 - 2(\$297.06) + \$268.28}{\$290.75 - \$268.25}$

$= 0.0391$

(d) Percentile co-efficient of skewness = $\frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}} = \frac{\$30\ \ .00 - 2(\$279.06) + \$258.\ \ 2}{\$30\ \ .00 - \$258.\ \ 2}$

$= 0.0233$

In (c) and (d) above, the co-efficients are positive, the distribution is skewed positively (i.e. to the right).

CLASS ASSESSMENT EXERCISE 4

1. The second moment about the mean of two distributions are 9 and 16 while the third moment about the mean are-8.1 and -12.8 respectively. Which distribution is more skewed to the left?

2. Given that the mean of a distribution is 80; the mode is 60 and standard deviation is 40. Find the co-efficient of skewness.
3. The third moment about the mean is 150.5. If the co-efficient of skewness of the distribution is 0.6, then:
   (i) Find the standard deviation.
   (ii) Calculate the variance.
4. The second moment about the mean of two distribution are 27 and 19 while the third moments about the mean are -6.2 and -10.4; which of the distribution is more skewed to the right?
5. Given Q_1 = 12.6, Q_3 = 50.9, P_{10} = 8.5 and P_{90} = 72.8, median = 35.5.
   (i) Find the percentile co-efficient of skewness.
   (ii) Obtain the quartile co-efficient of skewness.
   (iii) Interpret the answers obtained in (i) and (ii) above.

4.0 CONCLUSION

This unit has been able to expose students to the concept of skewness, measurement and interpretation of skewness. The relationship between frequency distribution, histogram and symmetrical distribution has equally be discussed.

5.0 SUMMARY

Skewness is the measure or degree of the departure of a distribution from symmetry or the measure of asymmetry of a distribution. A number of measures have been defined to measure the co-efficient of skewness of a distribution. Some of the measures include the Karl Pearson's first and second co-efficient of skewness, Bowley's quartile co-efficient of skewness, 10 – 90 percentile co-efficient of skewness and the moment co-efficient of skewness.

The skewness of a distribution can be described with respect to its direction and degree. The direction explains whether the distribution is positively or negatively skewed while the degree gives the numerical value (co-efficient of skewness) which ranges between 0 and 1.

6.0 TUTOR MARKED ASSIGNMENT

(1) The third moment about the mean of a distribution is 201.3. If the co-efficient of skewness of the distribution is 0.8, find:

   (a) The standard deviation

   (b) The variance

(2) Given that the mean of a distribution is 80, the mode is 60 and the standard deviation is 40. Find the co-efficient of skewness and interpret your result?

(3) Given that the lower and upper quartiles of a distribution whose median is 2.56, as 0.96 and 3.52 respectively, compute a co-efficient of skewness of the distribution.

(4) Find the third moment about the mean of a distribution with standard deviation 7.5 and the co-efficient of skewness is o.45.

(5) Given the number bags of rice sold per week at Duleko market by 50 traders as follows:

Bags of rice	1 – 5	6 – 10	11 – 15	16 - 20	21 - 25	26 – 30
No of sellers	7	10	16	9	5	3

a. Find the moment co-efficient of skewness using any measure of your choice?
b. Interpret the result obtained in (a) above.

(6) The table below shows the distribution of the maximum loads in short tons supported by certain cables produced by a company.

Maximum Load (short tons)	**Numbers of Cables**
9.3 – 9.7	2
9.8 – 10.2	5
10.3 – 10.7	12
10.8 – 11.2	17
11.3 – 11.7	14
11.8 – 12.2	6
12.3 – 12.7	3
12.8 – 13.2	1

a. Find the Pearson's (a) first and (b) second co-efficient of skewness for the distribution and account for the difference.
b. Find the (a) quartile and (b) percentile co-efficient of skewness for the distribution. Compare your results with those in (a) above and explain for the difference.

7.0 REFERENCE/FURTHER READING

Loto, M. A., Ademola, A. A. and Toluwase, J. A. (2008). Statistics Made Essay. Concept Publication Limited, Lagos, Nigeria.

Spiegel, M. R. and Stephen, L. J. (2007). Statistics. (Fourth Edition). Schaum's Outline, McGraw Hills, Singapore.

UNIT 3: KURTOSIS

Table of Contents

1.0 INTRODUCTION

Another method of describing a frequency distribution is known as kurtosis. This consists of specifying for the distribution, its degree of peakness or steepness or, to use a Greek word, kurtosis. Two distributions may have the same mean, and the same standard deviation. They may also be equally skewed, but one of them may be more peaked than the other.

This unit shall consider the meaning of kurtosis, types of kurtosis, measurement as well as interpretation of its co-efficients.

2.0 OBJECTIVES

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

(i) Define the term "kurtosis";
(ii) State and explain the forms or types of kurtosis and;
(iii) Identify the alternative measures of kurtosis, their computations as well as interpretations.

3.0 MAIN CONTENT

3.1 General Overview

Kurtosis is defined as the degree of peakness of a distribution when it is compared with a normal distribution. A distribution which is either highly peaked or low in peakness but is moderate is called mesokurtic distribution while a platykurtic distribution is less peaked than the normal or mesokurtic distribution. However, leptokurtic distribution has the highest peakness.

CLASS ASSESSMENT EXERCISE 1

Define the term "kurtosis".

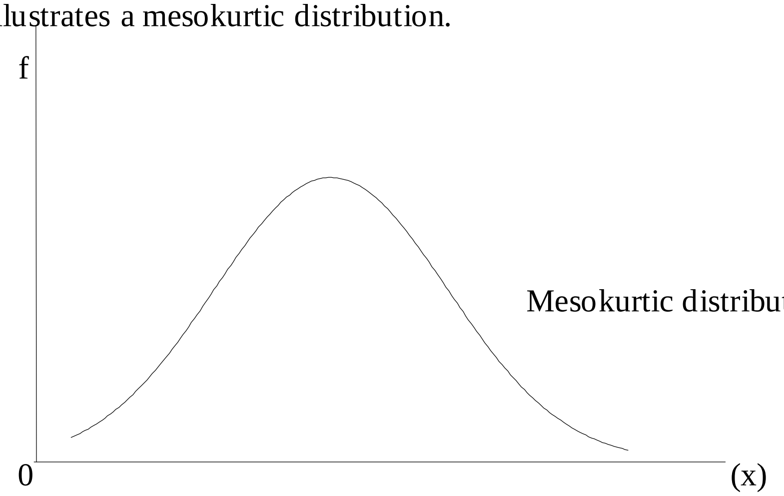
3.2 Types of Kurtosis

Kurtosis are of three types: -

(i) Mesokurtic distribution
(ii) Platykurtic distribution
(iii) Leptokurtic distribution

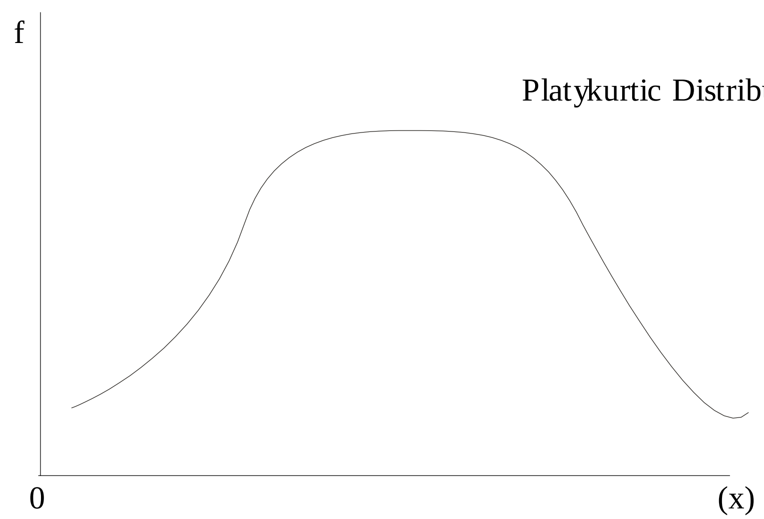
Mesokurtic Distribution

A normal distribution which is not very peaked or very low flat topped; is called mesokurtic distribution. It is a distribution which is neither highly peaked nor low in peakedness but moderate in nature. A mesokurtic distribution is a normal distribution that is symmetrical. The diagram below illustrates a mesokurtic distribution.



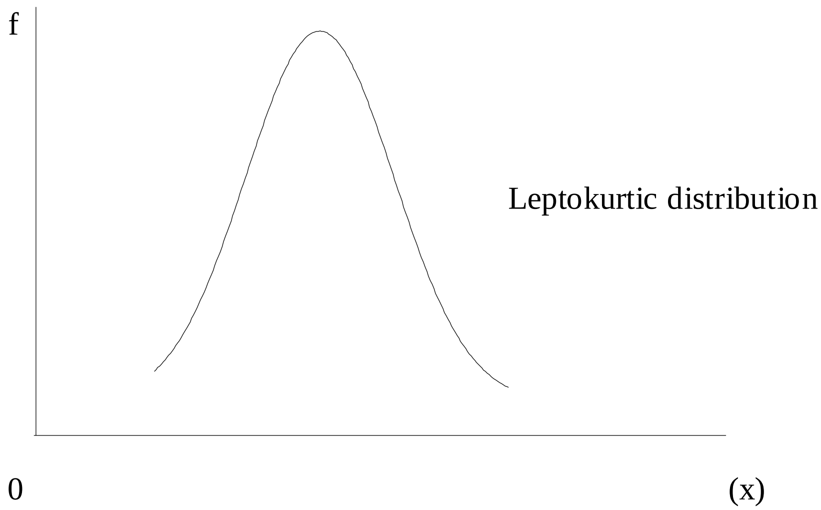
Platykurtic Distribution

A distribution which is flat topped is said to be leptokurtic. Such distributions are relatively less peaked than the normal curve or the mesokurtic distribution. The diagram below illustrates a platykurtic distribution.

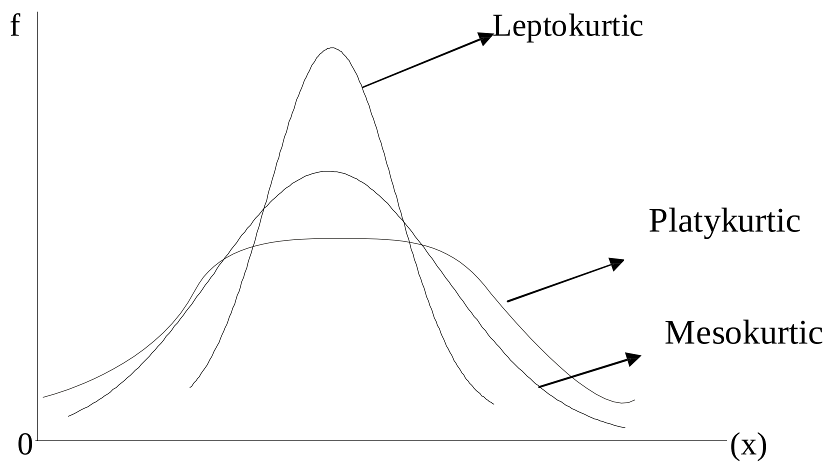


Leptokurtic Distribution

A leptokurtic distribution is the one that has the highest or greatest peakness among the three forms. Relative to platykurtic and mesokurtic distribution, a leptokurtic distribution has a greater steep-topped shape. The diagram below illustrates a leptokurtic distribution.



For quick relative comparison, the three types of kurtosis are illustrated in a two dimensional graph as follows:



CLASS ASSESSMENT EXERCISE

Outline and discuss with the use of appropriate graph, the three types of kurtosis.

3.3 Measurement and Interpretation of Kurtosis

Kurtosis, as a feature of a frequency distribution is measured using different approaches. Commonest among the measures are: -

(i) $e_2 = \frac{m^4}{S^4} - 3$ where e_2 = co-efficient of kurtosis, nt^4 = fourth moment about the mean and ^{4} is square of the variance. This answer is a pure number.

(ii) **Moment co-efficient of kurtosis: -** This measure made use of the fourth moment about the mean and the variance (i.e. 2nd moment about the mean). This is given as MCK (i.e. moment co-efficient of kurtosis),

MCK = $\mathbf{l}_4 = \frac{m_4}{S^4} = \frac{m_4}{(S^2)^2} = \frac{m_4}{m_2^2}$

For a normal distribution (mesokurtic distribution) $\frac{m_4}{S^4} = 3$. If we say that kurtosis coiuld be measured as $g_2 = \frac{m_4}{S^4} - 3$and we say that is a pure number, then, the first measurement for a normal distribution will turn out to be zero.

For leptokurtic distribution, $\frac{m_4}{S^4} > 3$ while;

For a platykurtic distribution, $\frac{m_4}{S^4} < 3$.

(iii) If $\frac{m_4}{S^4} = b_2$ of g_2 as we have in our first measurement of kurtosis, the co-efficient of kurtosis is defined as $b_2 - 3$ of $g_2 - 3$.

In this case, for platykurtic distribution, $b_2 - 3 < 0$ of $g_2 - 3 < 0$while for leptokurtic distribution, $b_2 - 3 > 0$ of $g_2 - 3 > 0$. It can also be given as: $3 + \frac{-6pq}{Npq}$

(iv) **Percentile co-efficient of kurtosis:** This is given as: K = $\frac{Q}{P_{90} - P_{10}}$ where Q is the semi-interquartile range (Q = $\frac{Q_3 - Q_1}{2}$) $\therefore$ K = $\frac{\frac{Q_3 - Q_1}{2}}{P_{90} - P_{10}}$

Or K = $\frac{Q_3 - Q_1}{2} \div p_{90} - p_{10}$

$= \frac{Q_3 - Q_1}{2} \times \frac{1}{P_{90} - P_{10}}$

$= \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$

For a mesokurtic distribution, k = 0.263

For a leptokurtic distribution, k > 0.263

For a platykurtic distribution, k < 0.263

Worked Examples

1. Given the fourth moment about the mean of a distribution as 130 and the variance as 4.3. Obtain the co-efficient of the kurtosis and comment on the result.

Solution

Here, we need to compute the co-efficient of kurtosis using the moment co-efficient of kurtosis; $\mathbf{l}_4 = \frac{m_4}{S^4}$

$^2 = 4.3$, $\ ^4 = (\ ^2)^2 = (4.3)^2 = 18.49$

$nt_4 = 130$

$\therefore \mathbf{l}_4 = \frac{m_4}{S^4} = \frac{30}{8.49} = 7.03$

The value is greater than 3. Therefore, the distribution is leptokurtic.

2. Twenty percent of electric bulbs manufactured by a company are found to be defective. Find the moment co-efficient of kurtosis for a distribution of defective bulbs in a total of 1,000 and interpret the result.

Solution

N = 1000, $P = \frac{20}{000} = 0.02$, q = 1 – 0.02 = 0.98

$$\text{Moment co-efficient of kurtosis} = 3 + \frac{-(6\times0.02\times0.98)}{000\times0.02\times0.98}$$
$$= 3 + \frac{-0.\quad 76}{96}$$
$$= 3 + \frac{0.8824}{96}$$
$$= 3 + 0.000450 = 3.00045$$

It is almost mesokurtic because the value is approximately equal to 3.

3. Given that $Q_1 = 56$, $Q_3 = 98$, $P_{90} = 110$ and $P_{10} = 42$, find the percentile co-efficient of kurtosis and comment on the peakness of the distribution.

Solution

Percentile co-efficient of kurtosis = $\frac{Q_3 - Q_1}{2(P_{90} - P_{10})} = \frac{98-56}{2(\quad 0-42)} = \frac{42}{36} = 0.309$

Since the value is greater than 0.263, the distribution is leptokurtic.

4. The standard deviation of a symmetrical distribution is 4.8. What should be the value of the fourth moment about the mean in order for the distribution to be:
   (i) Mesokurtic
   (ii) Platykurtic
   (iii) Leptokurtic

Solution

(i) Since s = 4.8, moment co-efficient of kurtosis $\frac{m_4}{S^4} = 3$, for mesokurtic distribution; $\frac{m_4}{(4.8)^4} = 3$

$M_4 = 3 \times (4.8)^4 = 1592.52$

Note: The value given is standard deviation not variance. Recall that variance is the square of standard deviation.

(ii) For platykurtic, $\frac{m_4}{S^4} < 3$

$\frac{m_4}{(4.8)^4} < 3 \therefore M_4 < 3 \times (4.8)^4$

$M_4 < 1592.52$

(iii) For leptokurtic distribution, $\frac{m_4}{S^4} < 3$

$\frac{m_4}{(4.8)^4} > 3 \therefore M_4 > 3 \times (4.8)^4$

$M_4 > 1592.52$

CLASS ASSESSMENT EXERCISE

1. If the fourth moment about the mean is 130 and the variance is 6.2. Compute and interpret the co-efficient of kurtosis for the distribution.
2. The fourth moment about the mean distributions and its standard deviation are 27 and 4.2. Obtain the co-efficient of kurtosis for the distribution.
3. Given that Q_1 = 12.6, Q_3 = 50.9, P_{10} = 8.5 and P_{90} = 72.8. Find the percentile co-efficient of kurtosis and comment on your result.

4.0 CONCLUSION

This unit has been able to examine the meaning and scope of kurtosis. Various types of kurtosis have been identified, defined and illustrated. Alternative measures of the degree of peakedness of a distribution (kurtosis) have also been examined. Some of these measures include the moment co-efficient of kurtosis and the percentile co-efficient of kurtosis. The major yardstick used in identifying the nature of the kurtosis from the co-efficient has been treated.

5.0 SUMMARY

Kurtosis measures the degree of peakness of a distribution. A distribution may be mesokurtic (neither highly peaked nor low in peakness) or platykurtic (less peak than the normal distribution i.e. mesokurtic) or leptokurtic (distribution with highest peakness). A number of measures can be used to obtain the co-efficient of kurtosis of a distribution. Prominent and common among them are moment co-efficient of kurtosis (MCK) and the percentile co-efficient of kurtosis. The value obtain from the co-efficient's computation determines the nature of kurtosis.

6.0 TUTOR MARKEDASIGNMENT

1. The variance of a symmetrical distribution is 36. What must be the value of the fourth moment about the mean in order that distribution to be
   (i) Mesokurtic
   (ii) Platykurtic
   (iii) Leptokurtic

2. Draw a diagram to illustrate
   (i) Leptokurtic distribution
   (ii) Mesokurtic distribution
   (iii) Platykurtic distribution

3. Given that the fourth moment about the mean is 199.3759 and the variance is 8.5275. Find and interpret the moment co-efficient of kurtosis

4. Given that Q_1 = 36, Q_3 = 80, P_{90} = 90, P_{10} = 36. Find the percentile co-efficient of kurtosis and comment on the peakness of the distribution.

5. Given the scores of students in an examination as follows:

Scores	1–5	6–10	11–15	16–20	21–25	26–30	31–35
Freq.	7	10	6	9	5	3	7

Find the moment co-efficient of kurtosis.

7.0 REFERENCES/FURTHER READING

Loto, M. A., Ademola, A. A. and Toluwase, J. A. (2008). Statistics Made Essay. Concept Publication Limited, Lagos, Nigeria.

Neil, A. Weiss (2008). Introductory Statistics (Eight Edition): Pearson International Edition, san Francisco Boston, New York.

Spiegel, M. R. and Stephen, L. J. (2007). Statistics. (Fourth Edition). Schaum's Outline, McGraw Hills, Singapore.

Module 5: Basic Statistical Measures of Estimates

Unit 1: Basic Concept in Probability

Unit 2: Use of Diagram in Probability

Unit 3: Elementary Probability Rules

Unit 4: Experimental Probability

Unit 5: Random Variable and Mathematics of Expectation

UNIT 1

BASIC CONCEPT IN PROBABILITY

Table of Contents

1.0 INTRODUCTION

The world we live is full of uncertainties. In our day to day activities, we cannot exactly predict the results of our efforts; weather cannot be predicted despite weather forecasts reported in the mass media, our lifes is equally unguaranteed. In order to cope with uncertainty of life, man make a lot of guesses, risks and gambles.

Business decisions are also based on uncertainty. Sales volume, profits, losses and other performance variables in business activities cannot be predicted. In the stock market, a stock dealer may decide to sell a particular share due to report that the price of the share may likely fall. These are just guesses based on facts, yet, there is no certainty. Inaccurate estimation by decision makers in business may lead to penalties ranging from more inconvenience to loss and bankruptcy. This calls for need to measure uncertainties accurately using the techniques of probability in order to take decisions that will minimise risk and danger.

Therefore, the issue of "uncertainty" utilises the branch of statistics called "probability" in order to arrive at the correct prediction about the population based on sample use. This unit, therefore, introduces you to the basic concepts of elementary probability theory.

2.0 OBJECTIVES

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

(i) Distinguish clearly between ‘events’, ‘experiments’, ‘sample space’ and ‘probability’.
(ii) Make a distinction between ‘mutually exclusive events’, ‘conditional events’ and ‘independent events’.

3.0 MAIN CONTENT

3.1 Events, Experiment, Outcome Sample Space and Probability

Events: A simple event is any single outcome from a probability experiment. Each simple event is denoted by e_i. Many different events can be associated with different experiments e.g. the event that the card selected from a pack of playing card is the king of hearts or the event that in a toss of die, a ‘3’ is obtained, or the event that in a toss of coins a ‘head’ is obtained etc. An event is therefore any collection of outcomes from a probability experiment. An event may consist of one or more simple events. Events are denoted using capital letters such as E. for example, event E = roll of even numbers from a toss of a die. Therefore, E = {2, 4, 6}.

An event can also be defined as a specific collection of sample points. Considering a simple experiment of tossing a die, the six sample points, S = {1, 2, 3, 4, 5, 6}. Since the die is balanced, each of the outcomes has equal chance of occurrence (i.e. $^1/_6$). An even number will occur if one of the sample points observe a ‘2’, or a ‘4’ or a ‘6’. A collection of sample points such as this is called an event. Therefore, the probability of obtaining an even number is $\frac{}{6}+\frac{}{6}+\frac{}{6}=\frac{3}{6}$ or ½. Hence, an event is a collection of outcomes for an experiment, that is, any subset of a sample space.

Experiment: In probability, an experiment is any process that can be repeated in which the results are uncertain. Probability experiment do not always produce the same results or outcome, so the result of any single trial of the experiment over many trials do produce regular patterns that enable us to predict with remarkable accuracy. For example, an insurance company cannot know ahead of time whether a specific 16-year old driver will be involved in an accident over the course of a year. However, based on historical records, the company can be fairly certain. An experiment is therefore an act or process of observation that leads to a single outcome that cannot be predicted with certainty.

Outcome: Probability is a measure of the likelihood of a random phenomenon or chance behaviour. Probability describes the long-term proportion with which a certain outcome will occur in situations with short term uncertainty. An outcome is defined as the result or set of results obtainable from an experiment. The long-term predictability of chance behaviour is best understood through a simple experiment. For instance, when a die is tossed 100 times, outcomes are expected, meaning that, each of the experiment of tossing a die gives an outcome.

Sample Space: The outcomes in an experimental probability cannot be decomposed into more basic parts. This is because observing the outcome of an experiment is similar to selecting the sample from a population. Therefore, the basic possible outcome of an experiment is called sample points. The sample space in an experiment is therefore defined as the collection of all its sample points. For instance, in the toss of a die; the sample space is S = {1, 2, 3, 4, 5, 6}. S_1{odd numbers: 1, 3, 5}, S_2{prime numbers: 2, 3, 5} etc are called sample points. A sample space is therefore the collection of all possible outcomes in an experiment.

Other examples include the sample space of results obtainable if a three coin are tossed simultaneously S = {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT} while the sample space of tossing two dice simultaneously can be obtained as follows:

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

Therefore, the sample space S = {(1,1) (1,2) (1,3) (1,4) (1,5) (1,6) (2,1) (2,2) (2,3) (2,4) (2,5) (2,6) (3,1) (3,2) (3,3) (3,4) (3,5) (3,6) (4,1) (4,2) (4,3) (4,4) (4,5) (4,6) (5,1) (5,2) (5,3) (5,4) (5,5) (5,6) (6,1) (6,2) (6,3) (6,4) (6,5) (6,6)}.

In a toss of four coins, the sample space can be obtained as: -

HHHH	HHTH	TTTT
HHHT	TTHH	TTTH
HHTH	HTHT	TTHT
HTHH	HTTH	THTT
THHH	THHT	HTTT

Hence, S = {HHHH, HHHT, HHTH, HTHH, THHH, THTH, TTHH, HTHH, HHTT, HTTH, THHT, TTTT, TTTH, TTHT, THTT, HTTT}

The number of all elements in a sample space is known as the sample size. In the previous sample spaces stated, the sample size of tossing three coins simultaneously is 8, that of tossing two dice simultaneously is 36 while that of tossing a fair coins four times or tossing four coins simultaneously is 16.

Probability: Probability is the ratio of the number of expected outcome to the number of all possible outcomes. It is the measure of the likelihood of the occurrence of an event. We cannot be absolutely certain of the occurrence of an event, and as such, we base most of social and economic decisions on likelihood. The probability that an event E occur is the number of ways in which all equally likely events (not bias), including E, occur. If E = an event, h = possible ways E can occur and n = total of all possible ways (sample size).

Then, probability of E: $P(E) = {}^{h}/_{n}$

The probability that E will not occur = $1 - P(E) = 1 - {}^{h}/_{n}$

$\therefore P(E) + P(E^1) = 1$ where $P(E^1)$ = probability that event E will not occur. The probability of any event is greater than or equal to zero and it is also less than or equal to 1. $\therefore 0 \leq p(E) \leq 1$.

When an event is certain, its probability is 1, e.g. in a bag containing 4 red balls, what is the probability of picking out a red ball at random. However, an event that is absolutely impossible has the probability of zero. For example, a bag consisting 4 green balls and 3 yellow balls; if a ball is picked at random, what is the probability that it is red? P(Red) = 0 because none of the ball is red hence picking a red ball from that bag is absolutely impossible and its probability is zero. The sum of the probability of simple events gives a total of 1 i.e. $P(e_1) + P(e_2) + P(e_3) + \dots + P(e_n) = 1$.

The above definition is referred to as the classical definition of probability. However, this definition has a limitation, it can be used when outcomes of events are equally likely to occur.

In some practical real life problems, not all outcomes are equally likely to occur. Some may have a higher degree of occurrence based on the previous information or data.

Example

Given that the scores recorded by an individual by tossing a coin 100 times is given as: -

Scores	1	2	3	4	5	6
Frequency	15	12	18	15	26	14

What is the probability that, the next dice he tosses give a score of 4. Note that in classical probability P(4) = 1/6 because all the scores are equally likely to occur. In this particular question, in which a table was given, P(4) = 15/100 and not 1/6, basically because of the previous event. This is called the Frequentist Definition of Probability in which the frequency approach (relative frequency) is applied.

However, there exists a modern probability theory (Axiomatic Definition of Probability) which tries to overcome the major limitation of the frequentist definition of probability. This limitation involves the non-existence of an actual limiting number in some cases when viewed theoretically. The modern probability theory tries to make up for this inadequacy by developing an axiomatic definition in which probability is regarded as an undefined concept.

CLASS ASSESSMENT EXERCISE 1

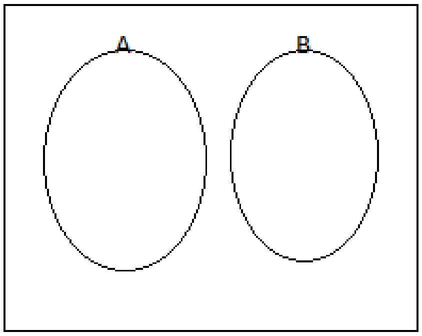
1. Define the following terms as related to probability: -
   a) Events
   b) Experiments
   c) Outcome
   d) Sample space
   e) Sample points

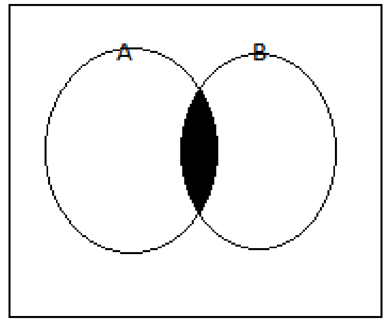
2. Define the concepts of probability from: -
   (i) Classical point of view;
   (ii) Frequentist point of view;
   (iii) Axiomatic point of view.

3.2 Mutually Exclusive, Conditional and Independent Events

Mutually Exclusive Events

Two events are said to be mutually exclusive if the occurrence of either excludes the possibility of the occurrence of other event. That is, either one or the other event but not both can occur at the same time. In other words, two or more events are mutually exclusive events if no two of them have outcome in common. If events A and B are mutually exclusive, the intersection of A and B is zero (null set).





Mutually exclusive event   Non-mutually exclusive event

Two or more events are mutually exclusive if the occurrence of one automatically implies the non occurrence of the other. For example, if E_1 represents the event of having a ‘head’ in a toss of a coin and E_2, the event of having a ‘tail’ in a toss of the coin. Then, E_1 and E_2 are mutually exclusive event since you cannot have both a head and a tail in a toss if the coin. If E_1 and E_2 are mutually exclusively, then $P(E_1 \cap E_2) = 0$.

Conditional Probability

If E_1 and E_2 are two events, the probability that E_2 occur given that E_1 has occur is denoted by $\mathrm{pf}\ {}^{E_2}/_{E}$ or Pr(E_2 given E_1) and is conditional probability of E_2 given that E_1 has occurred. $\mathrm{pf}\ {}^{E_2}/_{E}$ is read “probability of E_2 given E_1”. Hence, E_1 is called the given event. For example, the event that a student graduates from a tertiary institution is conditional to the event that he matriculates.

Example

1. In a toss of a fair die, what is the probability that:
   (i) a 5 is rolled, given that the die comes up odd.
   (ii) The die comes up odd, given that 5 is not rolled.

Solution

(i) $\mathrm{pf}\ {}^{\mathbf{fti}R}/_{o} = 1/3$. This is because there are three odd numbers in a die and 5 is put just one of them.

(ii) pf ^{o} / oa ftiR $= \Pr(1,3)$, Out of the five remaining possible outcomes (1, 2, 3, 4, 5, 6). Therefore, pf ^{o} / oa ftiR $= \frac{2}{5}$.

2. Given that joint probability distribution table is given as follows:

	Marital Status				
	Single M_1	**MarriedM_2**	**Widowed M_3**	**Divorced M_4**	**Total $P(S_i)$**
Male S_1	0.134	0.292	0.013	0.042	0.481
Female S_2	0.110	0.296	0.053	0.060	0.519
PM Total	0.244	0.588	0.102	0.102	1.000

Suppose one of the people is chosen at random;

(a) Determine the probability that the person is a selected is divorced, given that the person is a male.

(b) Determine the probability that the person selected is a male, given that is divorced.

Solution

(a) $p\ {}^{M_4}/\ \ = \frac{p(\ \ 1\ \ M_4)}{p(\ \ )} = \frac{P(S_1 \cap M_4)}{P(S_1)} = \frac{0.042}{0.48} = 0.087$ i.e. 8.7%

(0.087×100) male person are divorced.

(b) $p\ \ \ /_{M_4} = \frac{P(M_4\ \text{and}\ S_1)}{P(M_4)} = \frac{P(M_4 \cap S_1)}{P(M_4)} = \frac{0.042}{0.\ \ 02} = 0.412.$ i.e. 41.2% of the divorced people are male.

Note: $p\ {}^{E_2}/_{E} = \frac{P(E_1 \cap E_2)}{P(E_1)}$, where $p(E\ \ ) \neq 0$

Likewise, $p\ {}^{E}\ /_{E_2} = \frac{P(E_2 \cap E_1)}{P(E_2)}$, where $p(E\ \ ) \neq 0$

Independent Events

If two events are such that one has no effect on the other, then they are independent events. This implies that two events are said to be independent if the occurrence or non-occurrence of one event has no influence on the occurrence or non-occurrence of the other event. On the other hand, two events are dependent if the occurrence of one of the event in a probability experiment affects the probability of the other event. For example, in a toss of a coin and a die, the probability of obtaining ‘4’ on a die has nothing to do with the probability of obtaining a ‘head’ in the same simultaneous toss.

Event A is said to be independent of event B if $p\left({}^{A}/_{B}\right) = p(A)$. Then, the intersection of event A and event B is not equal to zero.

$p(A \cap B) \neq 0$

$p(A \cup B) = p(A) + p(B) - p(A \cap B)$

Example

1. If A and B are independent events, P(A) = 0.3, $p(A \cap B) = 0.16$, $p(A \cup B) = 0.92$. Find P(B).

Solution

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$
$$0.92 = 0.3 + p(B) - 0.16$$
$$0.92 = 0.3 - 0.16 + p(B)$$
$$p(B) = 0.92 - 0.3 + 0.16$$

2. Given K = event of a king selected in a pack of card (containing 52 playing cards) and
   H = event of a heart selected in the same pack.
   Determine whether event K is independent of event it or not.

Solution

$$p(K) = \frac{4}{52} \quad p(H) = \frac{13}{52} \quad p\left({}^{k}/_{H}\right) = \frac{p(K \cap H)}{p(H)} = \frac{\frac{4}{52} \times \frac{13}{52}}{\frac{13}{52}} = \frac{4}{52} = \frac{1}{13}$$

If the events are independent, $p\left({}^{k}/_{H}\right) = p(K) \therefore \frac{4}{52} = \frac{\ }{3}$. Then, event K is independent of event H.

CLASS ASSESSMENT EXERCISE

1. With appropriate examples and illustrations, distinguish clearly between mutually exclusive events and independent events.
2. When do you we say two events are having conditional probability.
3. Given the distribution of academic staff in a country as follows:

	Gender		
	Male G_1	**Female G_2**	**Total**
Professor R_1	21,224	3,194	24,418
Reader R_2	16,332	5,400	21,732
Senior Lecture R_3	25,888	14,491	40,379
Lecturer R_4	5,775	5,185	10,960
Assistant Lecturer R_5	781	723	1,504
Total	70,000	28,993	98,993

(a) Find $P(R_3)$

(b) Find p $^{3}/_{G}$

(c) Are events G1 and R3 independent? Explain your answer.

(d) Is the event that a person is female independent of the event that the person is an associate professor? Explain your answer?

4. Given that A = {odd number in a toss of a fair die}
B = {number less than 4 in a toss of fair die. Are A and B independent events?

5. Given that A and B are conditional events;

a. P(A) = 0.4, P(B) = 0.3, $p(A \cap B) = 0.06$; find

$p\ {}^{A}/_{B}$, $p\ {}^{B}/_{A}$, $p(K \cup H)$

b. If $p(A) = \frac{-}{4}, p(A \cap B) = \frac{3}{60}$; Ft $p\ {}^{B}/_{A}$

4.0 CONCLUSION

The unit has been able to introduce you to the basic concepts in probability.

5.0 SUMMARY

Probability is a branch of statistics that measures the degree of possibility of an event occurring in an experiment. Probability of an event (E) = $^{n}/_{h}$.

Where *n* is the number of possible way the event can occur and *h* is the total sample space.

Mutually exclusive events are events in which the occurrence or non-occurrence of one event precludes the occurrence or non-occurrence of the other event while independent events are events in which the occurrence of one event does not preclude the occurrence of the other event.

In computing the value of probability, it is pertinent to know the nature of the events involved.

6.0 TUTOR MARKED ASSIGNMENT

1. The table below shows the distribution of workers in an organization:

	Citizens (C)	Foreigners (F)
Administrative (A)	25	15
Sales (S)	30	40

Find (a) $p(F \cap C)$ (b) $p(\cap C)$ (c) $p(C/F)$ (d) $p(/A)$
(e) $p(C \cap F)$ (f) $p(/A)$

2. The probability that David and Paul would be present at a party are $^{1}/_{3}$ and $^{2}/_{5}$ respectively. Calculate the probability that;

   (i) Both of them would be present at the party.

   (ii) Only David would be present at the party.

   (iii) Only Paul would be present at the party.

   (iv) Neither of them would be present.

   (v) At least one of them would be present at the party.

   (vi) At most one of them would be present at the party.

7.0 REFERENCES/FURTHER READINGS

Loto, Marget A. et al (2008). Statistics Made Easy. Concept Publication Limited.

Sincich, McClaire and McClave, J. T. (2009). Statistics. Pearson International Edition. (11th Edition). Pearson Prentice Hall, Saddle River, USA.

Sullivan, Micjeal (2005). Fundamentals of Statistics. Pearson Prentice Hall, Upper Saddle River, New Jersey, USA.

Weiss, Neil A. (2008). Introductory Statistics. Pearson International Edition. (Eighth Edition). Montreal Munich Publishers, Singapore.

UNIT 2

USE OF DIAGRAM IN PROBABILITY

Table of Contents

1.0 INTRODUCTION

The use of diagrams and charts are equally used in probability to synthesise some practical problems. The commonest among such diagrams is the tree diagram and the Venn diagram. This unit introduces the student to the approach of using these diagrams in solving practical probability problems.

2.0 OBJECTIVES

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

(i) Solve practical probability problems using tree diagram;
(ii) Solve probability problems using Venn diagram.

3.0 MAIN CONTENT

3.1 Tree Diagram and Probability

A kind of three with branches indicating the probability of event or events at a particular point in time, may be used to compute probability values, this is commonly referred to as Tree Diagram.

Example

1. A sack contains six unbiased dice (each of six faces numbering 1 to 6). The dice are distinguished in six colours: yellow, white, red orange, green and blue. Use tree diagram to answer the following: -

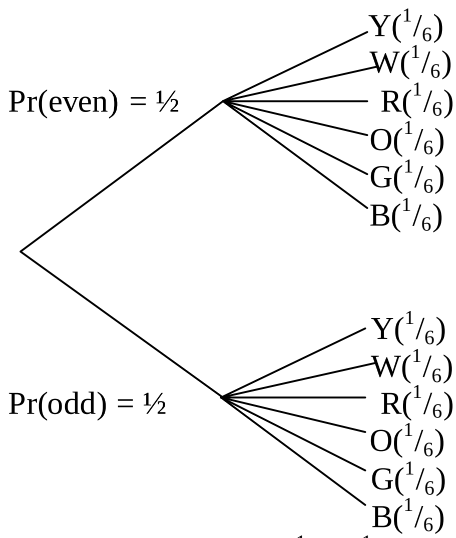
   (a) If a dice is picked at random, what is the probability that it is white and the score obtained from it is even?

(b) If a dice is picked at random, what is the probability that it is red with even score or a yellow with red score.

(c) If two dice are picked at random, what is the probability that they are green and white.

Solution

(1)

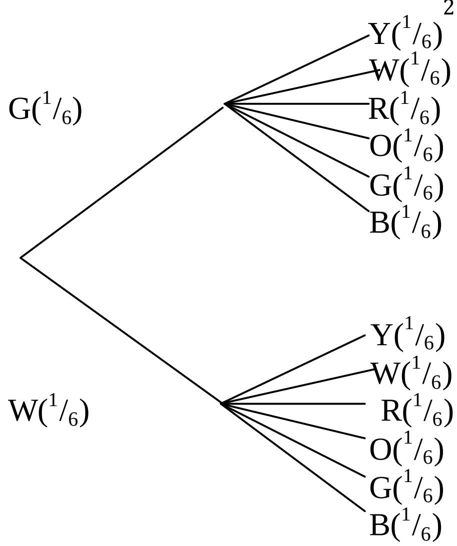


(a) ∴ Pr(even and yellow) = $\frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$

(b) Pr(even and red) or (odd and yellow) = $\frac{}{2} \times \frac{}{6} + \frac{}{2} \times \frac{}{6}$

$= \frac{}{2} + \frac{}{2} = \frac{2}{2} = \frac{}{6}$

(c)



Pr(Green and white) = Pr(green and white) or Pr(white and green)

$= \frac{}{6} \times \frac{}{6} + \frac{}{6} \times \frac{}{6}$

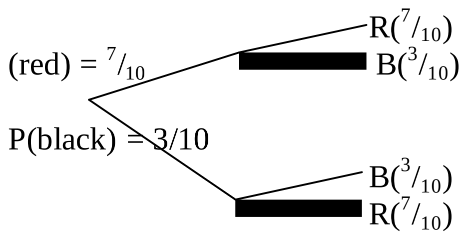
$= \frac{}{36} + \frac{}{36} = \frac{2}{36} = \frac{}{8}$

2. A bag contains 10 marbles, 7 are red and 3 are black. A marble is drawn at random and then replaced. Another draw is made

(i) What is the probability that both marbles drawn are red? (Hint: tree diagram)

(ii) What is the probability that two marbles selected are of the same colour?

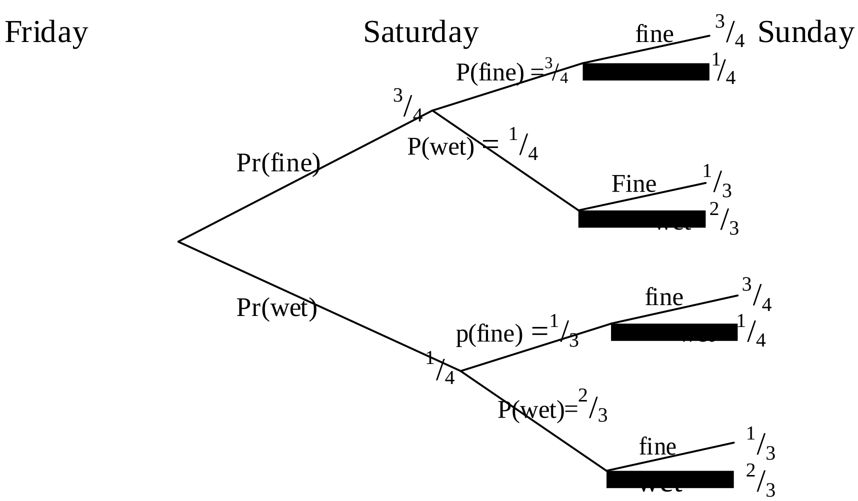
Solution



(i) Pr(both red) = P(R) and P(R) = $\frac{7}{0}\times\frac{7}{0}=\frac{49}{00}$

(ii) P(RR) or P(BB) = $\frac{7}{0}\times\frac{7}{0} + \frac{3}{0}\times\frac{3}{0}$

3. If it is five one day, the probability that it is fine the next day is ¾. If it wet one day, the probability that it is wet the next day is $^2/_3$. If it is fine today (Thursday), find the probability that it is fine on (i) Saturday (ii) Sunday; using tree diagram.



(i) The probability that it five on Saturday = $\frac{3}{4}\times\frac{3}{4}$ of $\frac{}{4}\times\frac{}{3}$

$= \frac{9}{6} + \frac{}{2} = \frac{3}{48}$

(ii) Probability that is fine on Sunday

$$= \frac{3}{4}\times\frac{3}{4}\times\frac{3}{4} \text{ of } \frac{3}{4}\times\frac{}{4}\times\frac{}{3} \text{ of } \frac{}{4}\times\frac{}{3}\times\frac{3}{4} \text{ of } \frac{}{4}\times\frac{2}{3}\times\frac{}{3} = \frac{347}{576}$$

CLASS ASSESSMENT EXERCISE 1

1. In particular filling station, 50% buy petrol, 305 buy diesel and the rest buy kerosene. If 40% of those who buy petrol fill their 50kg can, 605 of those who buy diesel fill their can and 70% of those who buy kerosene did not fill their tank. Assuming that a customer is picked at random, use tree diagram to find the probability that;

(a) He bought kerosene and fill the can.
(b) He bought either petrol or diesel but did not fill the can.
(c) He bought diesel and fill the can or he bought petrol but did not fill the tank.

2. Given the table below:

Region	Percentage of United State Population	Percentage of seniors
North West	19.0	13.8
Mid west	23.1	13.0
South	35.5	12.8
West	22.4	11.1
	100.0	

3. If a citizen is selected at random, use tree diagram to find the probability that he is a senior citizen from either the Midwest or the West.

3.2 Venn Diagram and Probability

The Venn diagram of the set theory is sometimes used in solving probability problems. Given a set of E_1 and E_2, the relationship between the two events depends on the nature of the events they are.

For mutually exclusive events, $p(E \enspace \cap E_2) = 0$

E **1** E_2 are mutually exclusive events, $E \cap E_2$ are mutually exclusive events.

$$p(E \enspace \cap E_2) = p(E \enspace) + p(E_2), \text{ since } p(E \enspace \cap E_2) = 0.$$

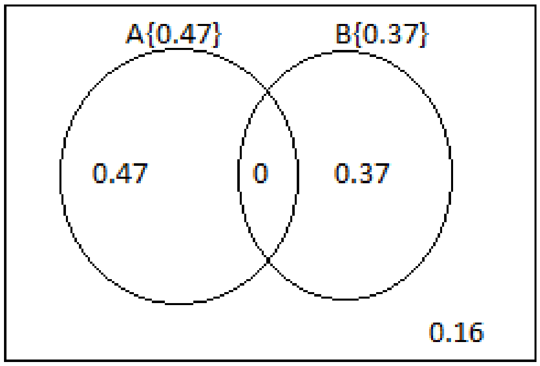
Note: Generally, for two events, $p(A \cup B) = p(A) + p(B) - p(A \cap B)$. But for mutually exclusive events, $p(A \cap B) = 0$. Since, it is not possible for two events to occur simultaneously.

Example1

Given that A and B are mutually exclusive events; P(A) = 0.47, P(B) = 0.33. Determine; (i) $p(A \cap B)$ (ii) $p(A \cup B)$ (iii) $p(A \ \cup B)$

(iv) $p(A \cup B)$

Solution



(i) $p(A \cap B) = 0$

(ii) $p(A \cup B) = p(A) + p(B)$

$= 0.47 + 0.37 = 0.84$

(iii) $p(A \ \cup B) = p(A \ ) + p(B)$

$= (1 - 0.47) + 0.37 = 0.90$

(iv) $p(A \cup B) \ = 1 - p(A \cup B)$

$= 1 - (0.47 + 0.37) = 1 - 0.84 = 0.16$

For independent events, E **1** E_2

E **1** E_2 are independent events

E **1** E_2 are independent events

E **1** E_2 are independent events

$p(E \ \cap \ E_2) = p(E \ ) \times p(E_2)$

$p(E \ \cup \ E_2) = p(E \ ) + p(E_2) - p(E \ \cap \ E_2)$

$= p(E \ ) + p(E_2) - [p(E \ ) \times p(E_2)\ ]$

Example II

Given that the probability that Ayo attends a party is independent of Bolu attending the same party. If the probability that Ayo attends is $^2/_3$ and the probability that Bolu attends is $^3/_5$;

(a) Find;

(i) The probability that both of them attend the party;

(ii) The probability that either of them attend the party.
(iii) The probability that none of them attend the party.
(iv) The probability that Ayo attends the party provided Bolu attended.

(b) Present the information, using Venn diagram.

(c) P(A ∪ B) where A = Probability that Ayo attends and B = Probability that Bolu attends.

Solution

P(A) = $^2/_3$

P(B) = $^3/_5$

a. (i) $p(A \cap B) = \frac{2}{3} \times \frac{3}{5} = \frac{2}{5}$

(ii) $p(A \cap B\ )$ of $p(A\ \ \cap B) = \frac{2}{3} \times\ 1 - \frac{3}{5}\ \ +\ \ 1 - \frac{2}{3}\ \times \frac{3}{5}$

$= \frac{2}{3} \times \frac{2}{5} + \frac{}{3} \times \frac{3}{5}$

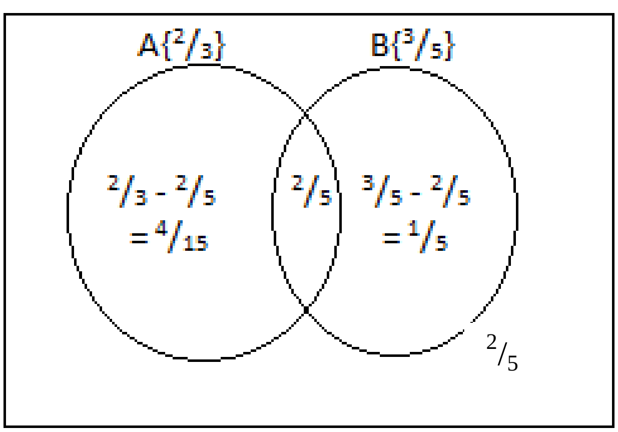
$= \frac{4}{5} + \frac{3}{5} = \frac{7}{5}$

(iii) $p(A\ \ \cap B\ ) =\ 1 - \frac{2}{3}\ \times\ 1 - \frac{3}{5}$

$= \frac{}{3} \times \frac{2}{5} = \frac{2}{5}$

(iv) $p\ ^A/_B = \frac{P(A \cap B)}{P(B)} = \frac{\frac{2}{3} \times \frac{3}{5}}{\frac{3}{5}} = \frac{2}{3}$

b.



c. $p(A \cup B) = p(A) + p(B) - p(A \cap B)$

$$= \frac{2}{3} + \frac{3}{5} - \frac{2}{5} = \frac{3}{5}$$

For conditional events, E_1 and E_2, the following holds;

(i) $p(E \cap E_2) = p(E\ ).\ p\ {}^{E_2}/_{E}$

(ii) $p(E \cup E_2) = p(E\ ) + p(E_2) - p(E\ ).\ p\ {}^{E_2}/_{E}$

From (i) above,

$$p\ {}^{E_2}/_{E} = \frac{P(E_1 \cap E_2)}{P(E_1)}; \text{ provided } p(E\ ) \neq 0$$

Likewise; $p\ {}^{E}/_{E_2} = \frac{P(E_1 \cap E_2)}{P(E_2)}$; provided $p(E_2) \neq 0$

Example III

If A and B are conditional events and P(A) = 0.40, P(B) = 0.30, P(A ∩ B) = 0.06. Find: (a) $p(A^c)$ (b) $p(B^c)$ (c) $p(A/B)$ (d) $p(B/A)$ (e) $p(A \cup B)$

Solution

(a) $p(A^c) = 1 - p(A) = 1 - 0.40 = 0.60$

(b) $p(B^c) = 1 - p(B) = 1 - 0.30 = 0.70$

(c) $p(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0.06}{0.30} = 0.2$

(d) $p(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{0.06}{0.40} = 0.15$

(e) $p(A \cup B) = p(A) + p(B) - p(A)\ .\ p(B/A)$

$= 0.4 + 0.3 - (0.4 \times 0.15)$

$= 0.7 - 0.06 = 0.64$

OR

$p(A \cup B) = p(A) + p(B) - p(A \cap B) = 0.4 + 0.3 - 0.06$

$= 0.64$

CLASS ASSESSMENT EXERCISE 2

1. A and B are two events such that P(A) = 0.3, P(A ∩ B) = 0.16, P(A ∪ B) = 0.92.

Required:

(a) Find P(B)

(b) Present the information using Venn diagram.

2. If A and B are mutually exclusive events. P(A) = 0.37, P(B) = 0.44. Find;

(a) $p(A\ )$ (b) $p(B\ )$ (c) $p(A \cup B)$ (d) $p(A \cap B)$

(e) $p(A \cap B\ )$ (f) $p(A\ \cup B)$

4.0 CONCLUSION

This unit has exposed you to the use of diagrams and charts in solving some practical and experimental probability problems.

5.0 SUMMARY

Tree diagram and Venn diagram are the most common approaches in solving probability problems with the aid of diagram. While tree diagram involve the drawing of a tree-like diagrams with branches illustrating the distribution of probability, the Venn diagram is the use of oval shapes drawn in a rectangular space. Both diagrams have the ability to solve various degree of practical probability problems with particular reference to the nature or type of the events involved.

6.0 TUTOR MARKED ASSIGNMENT

1. If B is the event that a car will have a faulty brake and T is the event that it will have bad tyre.

   (a) What type of events is B and T?

   (b) If P(B) = 0.23, P(T) = 0.24 and P(B ∪ T) = 0.38. Find P(B ∩ T). (Hint: Use Venn diagram).

2. Given that S = {1, 2, 3 … 9}

   A = {1, 3, 5, 7}
   B = {6, 7, 8, 9}
   C = {2, 4, 8}
   D = {1, 5, 9}

Find (a) $p(A\ \cap B)$ (b) $p(A\ \cap C) \cap B$ (c) $p(B\ \cup C)$

(d) $p(B\ \cup C) \cap D$ (e) $p(A\ \cap C)$ (f) $p(A\ \cap C) \cap D$

3. If S = $\{0/0 < x < 10\}$ M = $\{x/3 < x \leq 8\}$ and N = $\{x/5 < x < 10\}$

   Find (a) $p(M \cup N)$ (b) $p(M \cap N)$ (c) $p(M \cap N\ )$

   (d) $p(M\ \cup N)$

4. A man travels to work either by car or by train. There is a probability of 2/3 that he travels by train on any one day, there is probability of ¾ that

he will travel by car the next day. If he travels by car on any one day, there is a probability of $^5/_6$ that he will travel by train the next day. Find the probability that he travels:

(i) By car on Tuesday (ii)
By train on Tuesday
(iii) By car on Wednesday
(iv) By train on Wednesday
(Hint: Use tree diagram)

7.0 REFERENCES/FURTHER READINGS

Loto, Marget A. et al (2008). Statistics Made Easy. Concept Publication Limited.

Sincich, McClaire and McClave, J. T. (2009). Statistics. Pearson International Edition. (11th Edition). Pearson Prentice Hall, Saddle River, USA.

Sullivan, Micjeal (2005). Fundamentals of Statistics. Pearson Prentice Hall, Upper Saddle River, New Jersey, USA.

Weiss, Neil A. (2008). Introductory Statistics. Pearson International Edition. (Eighth Edition). Montreal Munich Publishers, Singapore.

UNIT 3

ELEMENTARY PROBABILITY RULES

Table of Contents	
1.0	Introduction
2.0	Objectives
3.0	Main Content
3.1	Basic Rules
3.2	Baye's Theorem
4.0	Conclusion
5.0	Summary
6.0	Tutor Marked Assignment
7.0	References/Further Reading

1.0 INTRODUCTION

Like many principles and topics in Mathematics and Statistics, probability is guided by some basic rules. This unit introduces you to the rules and how they are applied to practical and theoretical problems n experimental probabilities.

2.0 OBJECTIVES

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

(iii) State and apply the basic laws in solving probability problems.
(iv) Demonstrate the practicability of probability rules.

3.0 MAIN CONTENT

3.1 Basic Rules and their Application

(i) **Additive Rule**

$P(A \cup B) = P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
Note: P(A and B) = 0; for mutually exclusive events.

(ii) **Complement Rule**
$P(A')$ or $P(A^c)$ or $P(\bar{A}) = 1 - P(A)$.

(iii) **Multiplication Rule**
$P(A \cap B) = P(A \text{ and } B) = P(A) . P(B/A) \rightarrow$ general
$= P(A \text{ and } B) = P(A) . P(B) \rightarrow$ independent event

(iv) **Conditional Probability Rule**

$$P\ {}^{A}/_{B} = \frac{P(A \text{ and } B)}{P(A)} = \frac{P(A \text{ n } B)}{P(A)}$$

(v) **Rules of Total Probability**

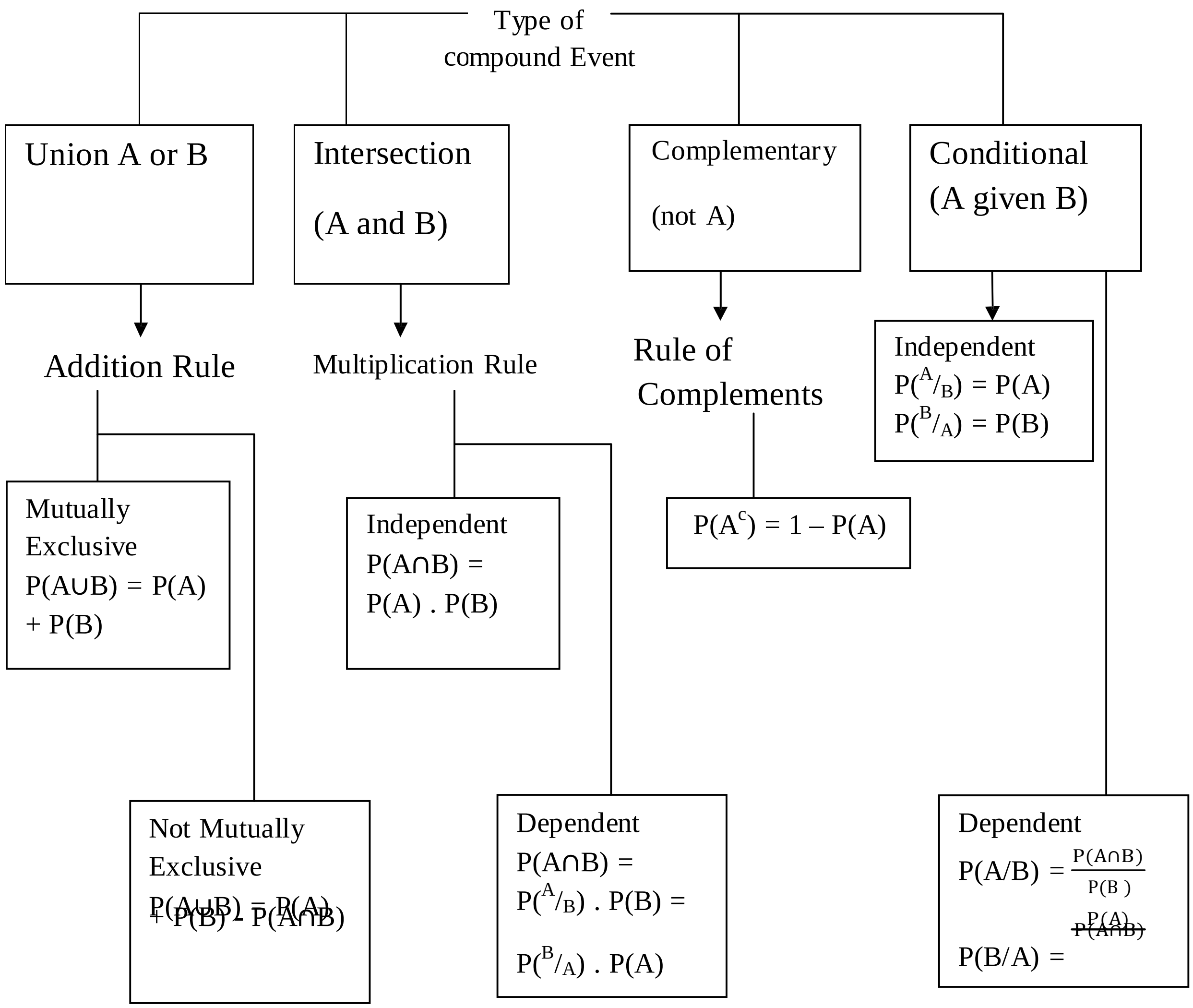
$$P(B) = \sum_{j=1}^{k} (P_j).P\left(\frac{B}{A}\right)$$

(vi) **Sum of Probability**

P(S) + P(f) = 1

Where, P(S) = probability of success
P(f) = Probability of failure

The chart below gives and adequate guide to selecting probability rules



CLASS ASSESSMENT EXERCISE 1

1. Distinguish clearly between mutually exclusive events and non-mutually exclusive events with respect to additional rule.
2. State the difference in the rule for multiplication between the independent events and dependent events.
3. What is the difference in the Conditional Rule when you compare independent events to dependent events?

3.2 Bayes Theorem

The theorem (otherwise known as inverse probability) was put forward by Thomas Bayes in his attempt to manipulate the formula for conditional probability in 1761. It is applicable in several practical experiments. In such experiments, it is possible that the outcome of an event may be caused by one or several possible factors.

It is appropriate when probability is obtained after the outcome of an experiment has been observed (Posteriori) rather than when the probability is known (Apriori).

Given k mutually exclusive and exhaustive events, $B_1, B_2, ..., B_k$ such that

$P(B_1) + P(B_2) + \dots + P(B_k) = 1$ and given an observed event A, it follows that

$$P\ {}^{B_i}/_{A} = P\ \frac{(B_i \cap A)}{p(A)}$$

$$= \frac{P(B_i).({}^{A}/_{B})}{P(B_i).P\ {}^{A}/_{B_i} + P(B_2).P\ {}^{A}/_{B_2} + \dots + P(B_k).P\ {}^{A}/_{B_k}}$$

This is known as the Bayes theorem.

Example 1

30% of the total output of a factory is produced in a workshop A while 70% is produced in workshop B. Also, 15 out of every 1000 components produced in A are defective and 8 out of every 1000 component produced in B are defective. If a component drawn at random from a mixture of the output of A and B is found to be defective, what is the probability that, it is from workshop A?

Solution

The probability that a component comes from A is $\frac{30}{00} = {}^{3}/_{10}$

The probability that it is defective given that it comes from A is

$P\left(^{D}/_{A}\right) = {^{15}/_{1000}} = 0.015$

Similarly P(B) = 70% = $^{70}/_{100} = 0.7$

Using Bayes theorem, the probability that the component comes from A and defective is

$$p\left(^{A}/_{D}\right) = \frac{P(A).P(^{D}/_{A})}{P(A).P(^{D}/_{A}) + P(B).P(^{D}/_{B})}$$

$$= \frac{0.3 \times 0.0\ \ 5}{0.3 \times 0.0\ \ 5 + 0.7 \times 0.008}$$

$$= \frac{0.0045}{0.0045 + 0.0056} = 0.45$$

Example 2

If a die is tossed 1 or 2 results a marble picked from bag a which contains 2 white and 3 red marbles; if 3, 4, or 5 results a marble picked from bag B which contains 1 white and 4 red balls while if 6 results a marble is picked from bag C which contains 3 white and 2 red balls. The experiment, when carried out results in a red ball being picked. Calculate the probability that the result of 6 is obtained from the tossed die.

Solution

If a result of 6 is obtained, the picking is done from bag C. So the probability of picking a ball from C, given that it is red is

$$P\left(^{C}/\right) = \frac{P(C)\ .P(^{R}/_{C})}{P(A).P(^{R}/_{A}) + P(B).P(^{R}/_{B}) + P(C).P(^{R}/_{C})}$$

Note: When 1 or 2 occurs, the picking is from bag A, So P(A) = $^{2}/_{6}$

since a die has 6 faces, when 3, 4, 5 occurs the picking is from bag B. So P(B) = $^{3}/_{6}$. When a 6 occurs, then picking is from C. So, P(C) = $^{1}/_{6}$.

Hence A, B, and C are mutually exclusive and exhaustive $\frac{2}{6} + \frac{3}{6} + \frac{}{6} = 1$

A contains 2 white and 3 red marbles and so, $P\ \ /_{A} = {^{3}/_{5}}$

B contains 1 white and 4 red marbles and so, $P\ \ /_{B} = {^{4}/_{5}}$

C contains 3 white and 2 red marbles and so, $P\ {}^{}/_{C} = {}^{2}/_{5}$

$$P\ {}^{C}/ = \frac{P(C).P({}^{R}/_{C})}{P(A).P({}^{R}/_{A}) + P(B).P({}^{R}/_{B}) + P(C).P({}^{R}/_{C})}$$

$$= \frac{\frac{1}{6}\times\frac{2}{5}}{\frac{2}{6}\times\frac{3}{5} + \frac{3}{6}\times\frac{4}{5} + \frac{1}{6}\times\frac{2}{5}}$$

$$= \frac{{}^{2}/_{30}}{{}^{6}/_{30} + {}^{12}/_{30} + {}^{2}/_{30}} = {}^{2}/_{30} \div {}^{20}/_{30}$$

$$= {}^{2}/_{30} \times {}^{30}/_{20} = {}^{1}/_{10} = 0.1$$

CLASS ASSESSMENT EXERCISE 2

Three types of capacitors A, B, C were installed in a certain make of transistor radios with 25% being of type A and 35% being of type B while the rest were from type C. The probabilities that A, B, and C are faulty in a year are 0.01, 0.015, and 0.02 respectively. If a capacitor fails within a year, what is the probability that it is of type A? (Hint: use Bayes theorem)

4.0 CONCLUSION

This unit has carefully exposed you to the basic rules in probability, including the Bayes theorem (inverse probability) and their basic application to problems.

5.0 SUMMARY

Rules used in probability theory include addition rule, complement rule, multiplication rule, conditional rule, etc. In applying these rules, the nature of the event has to be known. Bayes theorem is appropriate to obtain the probability of an experiment that has been observed (Posteriori) rather than the one that is already known (Apriori).

6.0 TUTOR MARKED ASSIGNMENT

1. In a toss of two dice, a sum less than 5 allows a person to pick balls form a bag containing 4 red ball and 2 green balls. A sum up to 5 but less than 10 allows him to pick ball from a bag containing 1 red ball and 3 green balls and a sum equals to 12 allows him to pick from a bag containing 3 red balls and 1 green ball. Given that a green ball is picked. Find the probability that it is picked by someone whose sum of scores on the dice is between 5 and 9. (Hint: use Bayes theorem)

2. Given the probability derived from a table as follows

$P(R_1) = 0.19$, $P(R_2) = 0.21$, $P(R_3) = 0.355$, $P(R_4) = 0.224$, P / = 0.138, P / 2 = 0.130, P / 3 = 0.128, and P / 4 = 0.11

Use Bayes theorem to find

(a) p 3/
(b) p /
(c) p 4/

7.0 REFERENCES/FURTHER READINGS

Loto, Marget A. et al (2008). Statistics Made Easy. Concept Publication Limited.

Sincich, McClaire and McClave, J. T. (2009). Statistics. Pearson International Edition. (11th Edition). Pearson Prentice Hall, Saddle River, USA.

Sullivan, Micjeal (2005). Fundamentals of Statistics. Pearson Prentice Hall, Upper Saddle River, New Jersey, USA.

Weiss, Neil A. (2008). Introductory Statistics. Pearson International Edition. (Eighth Edition). Montreal Munich Publishers, Singapore.

UNIT 4

EXPERIMENTAL PROBABILITY

Table of Contents

1.0 INTRODUCTION

Sometimes, there is a need to carry out probability experiments with the use of concrete objects such as coins, dice, playing cards, tabular data, etc. to be able to relate the theory of probability to everyday activities. This unit exposes you to the techniques involved in experimental probability with the use of concrete materials as well as the differences in computation of probability when applied.

2.0 OBJECTIVES

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

(i) Compute the values of probability given a set of information relating to concrete objects;
(ii) Compute probability values with different forms of selection.
(iii) Use tabular data to solve some probability problems.

3.0 MAIN CONTENT

3.1 Materials Used in Experimental Probability

(a) **Coin: -** A coin has two sides namely head (H) and a tail (T). Possible outcome for tossing variety of numbers of coins is given as:-

One coin: H, T

Two coins: HH, HT, TH, TT

Three coins: HHH, HHT, HTH, THH, TTT, TTH, THT, HTT

Example

1. In a toss of three coins, what is the probability of obtaining at least a tail?

Solution

P(at least a tail) = P(a tail) or P(2 tail) or P(3 tail)

= P(1T) + P(2T) + P(3T)

$= {}^{3}/_{8} + {}^{3}/_{8} + {}^{1}/_{8} = {}^{7}/_{8}$

2. Five coins are tossed, what is the probability that they all show the same faces?

Solution

To show same faces, we obtain {HHHHH o**f**TTTTT}, which is 2 possible outcome out of the 25(5^2) possible outcomes.

∴P(obtaining same faces) = ${}^{2}/_{32} = {}^{1}/_{16}$ of 0.0625

3. Six coins are tossed. Find the probability of obtaining a head.

Solution

To obtain one head, the outcomes are:
{HTTTTT, THTTTT, TTHTTT, TTTHTT, TTTTHT, TTTTTH}, that is, 6 probable events. For 6 coins, the total possible events is $2^6 = 64$

∴ P(one head) = ${}^{6}/_{64} = {}^{3}/_{32}$

(b) **Dice:** A common fair die has 6 faces; each of them has equal chance of occurring when it is tossed. Therefore

P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = ${}^{1}/_{6}$

When two dice are tossed the possible sum and combination are given as follows:-

Combination

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

Sum

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Example

1. In a toss of 2 dice what is the probability of obtaining

(a) a total sum of 10
(b) a sum less than 8
(c) a prime number sum
(d) same score
(e) one of the score being '3'

Solution

(a) P(sum of 10) = $^{3}/_{36}$

(b) P(sum less than 8) = $^{21}/_{36}$

(c) P(a prime number sum)

P(sum of 2) or P(sum of 3) or P(sum of 5) or P(sum of 7) or P(sum of 11)

$= {}^{1}/_{36} + {}^{2}/_{36} + {}^{4}/_{36} + {}^{6}/_{36} + {}^{2}/_{36}$

$= {}^{15}/_{36}$

(d) P(same score) = P$\{(1,1) \text{ of } (2,2) \text{ of } (3,3) \text{ o}\mathbf{f} (4,4) \text{ of } (5,5) \text{ of } (6,6)\}$

$= {}^{6}/_{36} = {}^{1}/_{6}$

(e) P(one of the score being a '3'). From the combination table = $^{10}/_{36}$.

Note: only one of the die being '3', not both of them being '3'.

2. Two dice and two coins are tossed at the same time, what is the probability of obtaining a tail and a sum of 9?

Solution

Two coins: HH, HT, TH, TT

P(T) = $^2/_4 = {^1/_2}$

P(sum of 9) = $^4/_{36}$

P(Tail and sum of 9) = $^1/_2 \times {^4/_{36}} = {^1/_{18}}$

3. Three dice are tossed, what is the probability of obtaining the same score throughout?

Solution

The total possibilities for tossing three dice is $6^3 = 216$

The same scores: {(1,1,1), (2,2,2), (3,3,3), (4,4,4), (5,5,5), (6.6,6)}

P(same score) = $^6/_{216} = {^1/_{36}}$

(c) **Playing Card:** A pack of playing card is 52 in number (the JOKER cards exclusive). There are 13 spades, 13 clubs, 13 hearts, and 13 diamonds, each having a king, an ace, a queen, a jack and number 2 to 10. Thus, we have "ace of spades", queen of hearts", "king of clubs" etc. The distribution of the contents of the pack of playing cards are shown below:

	King	Queen	Ace	Jack	2	3	4	5	6	7	8	9	10	Total
Club	1	1	1	1	1	1	1	1	1	1	1	1	1	13
Diamonds	1	1	1	1	1	1	1	1	1	1	1	1	1	13
Hearts	1	1	1	1	1	1	1	1	1	1	1	1	1	13
Spades	1	1	1	1	1	1	1	1	1	1	1	1	1	13
Total	4	4	4	4	4	4	4	4	4	4	4	4	4	52

Example

1. A child will be given a price if he draws either a king or a 9 in a pack of playing card, find the probability that he wins.

Solution

Check the workings

P(a 9) and P(a king) = $^4/_{36} \times {^4/_{36}}$

$= {^1/_9} \times {^1/_9} = {^1/_{81}}$

2. If a pack of a playing card properly shuffled, what is the probability of picking a Diamond card with and odd number?

Solution

P(Diamond) and P(odd number) = $^{13}/_{52} \times$ [p(3) of p(5) of p(7) of p(9)]

$= {^{13}/_{52}} \times {^4/_{52}} + {^4/_{52}} + {^4/_{52}} + {^4/_{52}}$

$= {^{13}/_{52}} \times {^{16}/_{52}} = {^1/_4} \times {^4/_{13}} = {^1/_{13}}$

CLASS ASSESSMENT EXERCISE 1

1. In a toss of four coins, determine the probability of obtaining
   (a) two heads
   (b) at most 3 tails
   (c) no head
   (d) at least 1 tail
2. In a toss of 2 dice, what is the probability of obtaining
   (a) Two prime numbers
   (b) A sum greater than 7
   (c) One of the scores being a prime number.
   (d) At least one of the scores being a '5'
3. In a pack of playing card,
   (a) If a card is selected at random, what is the probability that;
      (i) It is either a heart of a diamond
      (ii) It is a spade and a jack
   (b) If two cards are selected at random (with replacement), find the probability that;
      (i) Both of them are 'ace'
      (ii) One is a queen and the other is a jack.

3.2 SELECTION MODE IN PROBABILITY

Selection process in probability is usually in two folds – selection with replacement and selection without replacement. If we are given a bag containing ball of the same sizes and weights but different colours, then selection can be done either with replacement or without replacement. If there is no replacement, the number will be reducing (both the type being selected and the total sample space). However, if the selection is with replacement, the number is kept constant.

Example

1. A grocer bag contains 25 red and 35 green apples. Two apples are picked from the bag with replacement. Find the probability that
   (a) They are both red
   (b) They are both green
   (c) They are of the same colour.

Solution

(a) P(they are both red) = $P(R_1)$ and $P(R_2)$

$$= {}^{25}/_{60} \times {}^{25}/_{60}$$

$$= {}^{625}/_{3600} = 0.17$$

(b) P(they are both green) = $P(G_1)$ and $P(G_2)$

$$= {}^{35}/_{60} \times {}^{35}/_{60}$$

$$= {}^{1225}/_{3600} = 0.34$$

(c) P(they are of different colour) = P(G & R) or P(R & G)

$$= {}^{25}/_{60} \times {}^{35}/_{60} + {}^{35}/_{60} \times {}^{25}/_{60}$$

$$= {}^{875}/_{3600} + {}^{875}/_{3600} = {}^{1750}/_{3600}$$

$$= 0.486$$

2. In a bag containing 5 green balls and 7 yellow balls. If 2 balls are picked at random without replacement, find the probability that;
   (a) They are of the same colour
   (b) They are of different colour
   (c) At least one of them is green
   (d) None of them is yellow
   (e) The first is yellow and the second is green.

Solution

(a) P(same colour = P(G_1 and G_2) or P(Y_1 and Y_2)

$$= {}^{5}/_{12} \times {}^{4}/_{11} + {}^{7}/_{12} \times {}^{6}/_{11}$$

$$= {}^{20}/_{132} + {}^{42}/_{132} = {}^{65}/_{132}$$

$$= 0.47$$

Note: You observe that the value of the denominator decreases because as the ball is picked without replacement, the quantity of balls left in the bag decreases as selection continues.

(b) P(different colour) = P(G & Y) or P(Y & G)

$$= {}^{5}/_{12} \times {}^{7}/_{11} + {}^{7}/_{12} \times {}^{5}/_{11}$$

$$= {}^{35}/_{132} + {}^{35}/_{132} = {}^{70}/_{132}$$

$$= 0.53$$

(c) P(at least one is green) = P(G & Y) or P(Y & G) or P(G & G)

$$= {}^{5}/_{12} \times {}^{7}/_{11} + {}^{7}/_{12} \times {}^{5}/_{11} + {}^{5}/_{12} \times {}^{4}/_{11}$$

$$= {}^{35}/_{132} + {}^{35}/_{132} + {}^{20}/_{132} = {}^{90}/_{132}$$

$$= 0.682$$

(d) P(none is yellow) = P(Y') . P(Y') = P(G) and P(G)

$$= {}^{5}/_{12} \times {}^{4}/_{11}$$

$$= {}^{20}/_{132} = 0.15$$

(e) P(the first is yellow and the second green) = P(Y) and P(G)

$$= {}^{7}/_{12} \times {}^{5}/_{11}$$

$$= {}^{35}/_{132} = 0.265$$

CLASS ASSESSMENT 2

1. A bag contains 6 black, 4 red and 8 white balls. Three balls are picked at random from the bag. Find the probability that
   (a) They are of the same colour
   (b) They are of different colours

(c) They are picked in the order:- white, red, black, if there is no replacement.

2. A bag contains 3 green balls, 5 white balls and 7 yellow balls. Two balls are picked at random with replacement. Find the probability that
   (a) They are all yellow.
   (b) They are of the same colour.
   (c) At least one is yellow.

3.3 COMPUTING PROBABILITY FROM TABULAR DATA

Most times, table(s) may be given and some simple question sets of probability followed. In such problems, there may be a need to compute the horizontal and vertical totals before the question could be appropriately answered.

Example 1

The table below shows the number of students from three different schools who registered from four subjects in the last SSCE.

	French	Fine Art	Music	Hausa
School A	10	15	8	2
School B	14	20	10	6
School C	16	18	12	6

If a student is picked at random, find the probability that

(i) He/she is from school A.
(ii) He/she is from school C.
(iii) He/she is offering Fine Art from school B.
(iv) He/she is not offering Music and is not from school A.
(v) He/she is offering Hausa from school A.

Solution

	French	Fine Art	Music	Hausa	Total
School A	10	15	8	2	35
School B	14	20	10	6	50
School C	16	18	12	6	52
Total	40	53	30	14	137

(i) P(school A) = $^{35}/_{137} = 0.255$

(ii) P(school C and Hausa) = $^{6}/_{52}$ (because out of 52 student from school C, only 6 offer Hausa)

(iii) P(Fine Art and School B) = $^{20}/_{53}$ (because 53 students offer Fine art, out of which 20 attend school B)

(iv) P(he/she is not offering Music and is not from school A)

No of students who are not offering Music = 137 – No that offer Music

= 137 – 30 = 107

Out of this, 14 + 16 + 20 + 18 + 6 + 6 is not from School A i.e. 80 are not from School A

∴ The required probability = $^{80}/_{107} = 0.75$

(v) P(he/she is offering Hausa from school A)

14 students are offering Hausa, out of which 2 are from School A. The required probability = $^{2}/_{14} = 0.143$

(2) The scores of 40 students in an examination is given as follows

Scores	1-10	11-20	21-30	31-40	41-50
Frequency	9	6	7	8	10

(a) If a student is chosen at random, what is the probability that

(i) He scores below 21

(ii) He scores at least 31

(iii) He scores at most 30

(iv) He passes if the pass mark is 21.

(b) If two students are picked at random with replacement, find the probability that;

(i) They both failed if the pass mark is 21

(ii) One passed and the other failed if pass mark is 21

(iii) Both of them score above 30.

Solution

(a)(i) P(score below 21) = $^{9}/_{40} + ^{6}/_{40} = ^{15}/_{40}$

$= ^{3}/_{8}$ or 0.375

(ii) P(score of at least 30) = $^{8}/_{40} + ^{10}/_{40} = ^{18}/_{40} = 0.45$

(iii) P(scoring at most 30) = $^{9}/_{40} + {^{6}/_{40}} + {^{7}/_{40}} = {^{22}/_{40}} = 0.55$

(iv) P(passing is pass mark is 21) = $^{7}/_{40} + {^{8}/_{40}} + {^{10}/_{40}} = {^{25}/_{40}} =$ 0.625

(b)(i) P(fail) = $^{9}/_{40} + {^{6}/_{40}} = {^{15}/_{40}}$

$\therefore$ P(both fail) = $^{15}/_{40} \times {^{15}/_{40}} = 0.14$

(ii) P(fail) and P(pass) or P(pass) or P(fail)

$= {^{15}/_{40}} \times {^{25}/_{40}} + {^{25}/_{40}} \times {^{15}/_{40}}$

$= {^{375}/_{1600}} + {^{375}/_{1600}} = {^{750}/_{1600}}$

= 0.47

(iii) P(scoring above 30) = $^{8}/_{40} + {^{10}/_{40}} = {^{18}/_{40}}$

$\therefore$ P(both scoring above 30) = $^{18}/_{40} \times {^{18}/_{40}} = {^{324}/_{1600}}$

= 0.2

CLASS ASSESSMENT EXERCISE 3

People of 100 chronic and 400 non chronic carrier contain antigen reveal the following groups.

	O	A	B	AB
Carrier	50	35	10	5
Non Carrier	215	135	40	10

Find the probability that a person picked at random is: -

(a) A carrier;
(b) A carrier from group A.;
(c) A non-carrier or has blood group AB; and
(d) A non-carrier with blood group A.

4.0 CONCLUSION

This unit has been able to expose you to the basic application of probability rules in solving day to day practical problems using concrete and common objects around you. The use of tabular data has also been identified as an instrument of solving probability problems. The basic differences in selecting objects with replacement and without replacement have equally been identified.

5.0 SUMMARY

Various concrete objects that may be used to demonstrate the practicability of probability theory include coins, dice and playing cards. Probability values depend largely on the mode of selection as well as the type of events being considered.

6.0 TUTOR MARKED ASSIGNMENT

1. If 2 dice are tossed simultaneously, what is the probability of obtaining: -

   (a) A sum of 7;
   (b) A sum less than 10;
   (c) An even number sum;
   (d) Same scores on the dice; and
   (e) One of the scores being an "odd number".

2. A die and a coin are tossed simultaneously.

   Required:

   (a) The sample space.
   (b) What is the sample size?
   (c) What is the probability of obtaining: -
      (i) An odd score and a tail
      (ii) A score less than 3 and a head
      (iii) A prime number score and a head

3. A sack contains 4 red marbles, 5 yellow marble and the rest green. If there are 12 marbles in the sack and 2 are picked at random without replacement, find the probability that;

   (a) They are of the same colour.
   (b) They are of different colours
   (c) At least one of them is yellow.

4(a) Out of every 100 cheques handled by a cashier, 70 are honoured and the rest dishonoured, 20 are marked for representation and 10 are returned to the drawer. If two cheques are presented to the cashier, what is the probability that;

   (i) the two cheques would be honoured
   (ii) only one of the cheques would be honoured
   (iii) the two cheques would be dishonoured.

(b) Mr. Johnson presents a cheque to the cashier but the cheque is not honoured (dishonoured). Find the probability that it will be requested for representation.

5. The following table shows the blood groups and severity of a certain condition on a sample of 100 people.

Blood groups	Severity of condition		
A	300	70	10
B	200	30	5
AB	80	8	4
O	250	32	11

Calculate the probability that

(a) A person selected has blood group A
(b) A person with blood group O exhibits severe condition
(c) A person recognised to be free from the condition has blood group AB.
(d) A person has a mild condition or belongs to a blood group B

6. Two salesmen market a products using two new method A and B. if method A is used, probability of success is $^5/_6$ while if method B is used, the probability of success is $^1/_2$.

(a) Are methods mutually exclusive or independent events? How do you know?
(b) Thirty percent of the salesmen used method A while the remainder use B. If one of the sale successes is chosen from the whole list of successes, what is the probability that it was the result of method A?

7. If $P(E_1) = {}^1/_5$, $P(E_2) = {}^3/_4$, $P(E_1 n E_2) = {}^2/_{15}$

(a) Are events E_1 and E_2 independent or mutually exclusive? How do you know?
(b) Find (i) $P(E^c)$ (ii) $P(E_2^c)$ (iii) $P(E^c \;\; E^c)$

(iv) $P \; {}^{E_2}/_{E}$ (v) $P(E \; uE_2)$ (vi) $P(E^c n E_2^c)$

(vii) $P \; {}^{E_2}/_{E}$

8. The probabilities that Amaka and Abu will pass statistics examination are $^2/_3$ and $^4/_7$ respectively, find the probability that

(a) both of them pass the examination
(b) one of them pass the examination
(c) none of them pass the examination
(d) at least one of them pass the examination

7.0 REFERENCES/FURTHER READINGS

Loto, Marget A. et al (2008). Statistics Made Easy. Concept Publication Limited.

Sincich, McClaire and McClave, J. T. (2009). Statistics. Pearson International Edition. (11th Edition). Pearson Prentice Hall, Saddle River, USA.

Sullivan, Micjeal (2005). Fundamentals of Statistics. Pearson Prentice Hall, Upper Saddle River, New Jersey, USA.

Weiss, Neil A. (2008). Introductory Statistics. Pearson International Edition. (Eighth Edition). Montreal Munich Publishers, Singapore.

UNIT 5

RANDOM VARIABLE AND MATHEMATICS OF EXPECTATIONS

Table of Contents	

1.0 INTRODUCTION

In the last two units, we dealt with the probability distributions in which the variables have finite populations or sample space. However, many variables are not of this type: the number of people waiting in a line at a bank, outcome of tossing a coin or a die for a long period of time, the lifetime of an automobile, the weight of a new born baby, to mention a few.

Probability theory enables us to extend concepts that apply to variables of finite populations to other types of variables (like the ones mention above). In so doing, we are led to the notion of random variable and its probability distribution. The term random variable is more meaningful than the term variable alone, because the adjective random indicates that the experiment may result in one of the several possible values of the variable. For instance, in the toss of two coins, the possible number of heads obtainable is 0, 1, 2; according to the random outcomes of the experiment: HH, HT, TH, and TT.

This unit shall expose you to the meaning and scope of random variables, construction of probability histogram, random variable distribution, their mean, variance and standard deviation. The concept of mathematical expectation with respect to simple probability distribution shall also be treated.

2.0 OBJECTIVES

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

(i) Define and explain the concept of random variables.
(ii) Describe random variable distribution with respect to their probabilities, mean, variance, and standard deviation; and
(iii) Carry out simple mathematical problems involving expectation.

3.0 MAIN CONTENT

3.1 Meaning and Scope of Random Variables

A random variable is a variable that assumes numerical values associated with the random outcomes of an experiment, where one (and only one) numerical value is assigned to each sample point. It can also be described as a numerical measure of the outcome of a probability experiment, so its value is determined by chance. Random variables are denoted using letters such as x. for example, if an experiment is conducted in which a single die is cast, then x represents the number of pips showing on the die and the possible values of x are x = 1, 2, 3, 4, 5, 6.

A random variable is a quantitative variable or otherwise whose value depends on chance. This is because there are some random variables that are not quantitative; such variables include the sum of the dice when a pair of dice are rolled, the sum of puppies in a litter, the return of an investment, the lifetime of a flashlight battery, etc.

There are two types of random variables namely:-

(i) Discrete random variable.
(ii) Continuous random variable.

A discrete random variable is a random variable that has either a finite number of possible values or a countable number of possible values. That is, it is a random variable where all possible values can be listed. For instance, in a toss of two coins, all the possible outcomes can be listed as {HH, HT, TH, TT }. Hence, a random variable that can assume a countable number of values are said to be discrete.

A continuous random variable on the other hand is a random variable that has an infinite number of possible values that is not countable. Examples include the length of time (in minute) it takes a student to complete a one-hour exam: $0 \leq x \leq 60$, the depth (in feet) at which a successful oil-drilling venture first strikes oil $0 \leq x \leq C$, where C is the maximum depth obtainable. This module discusses the discrete random variables and their probability distribution. The continuous random variables and their probability distribution shall be discussed in Statistics courses.

CLASS ASSESSMENT EXERCISE 1

(1) Explain the term "Random Variable"
(2) Identify and explain the types of random variables.

3.2 Probability Histogram: A probability histogram is a histogram in which the horizontal axis corresponds to the value of the random variable and the vertical axis represents the probability of that value of the random variable. It is a graph of the probability distribution that displays the possible values of a discrete random variable on the horizontal axis. The distribution of each value is represented by a vertical bar whose height equals the probability.

Example 1

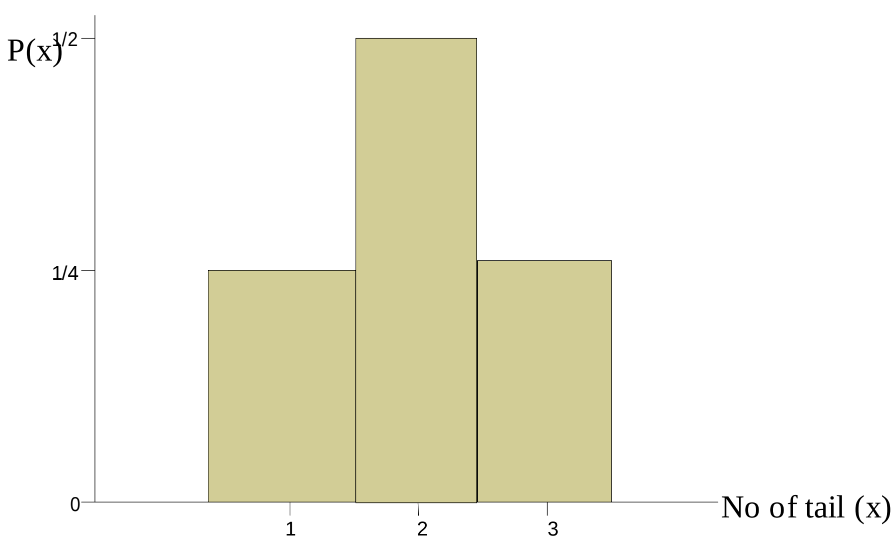
Given that the probability of obtaining a tail in a toss coins is P(x), where x is the number of tail(s) obtained.
Required:
(i) The sample space
(ii) The probability distribution of the random variable (x)
(iii) The probability histogram

Solution

(i) The number pace = {HH, HT, Th, TT}
(ii)

X(Number of Tail)	0	1	2
P(X)	¼	2/4 =1/2	¼

(iii)



2. Given that $f(x) = ({}^4C_x) \cdot \frac{}{2^4}$ where x = 0, 1, 2, 3 and 4

(a) Find f(0), f(1), f(2), f(3) and f(4).

(b) Present the information using histogram distribution

Solution

(a) $f(0) = {}^4C_0 \cdot \frac{}{2^4} = \frac{4!}{0!(4-0)!} \times \frac{}{6} = \frac{4!}{0!4!} \times \frac{}{6}$

Recall 0! = 1

$f(o) = \frac{4!}{4!} \times \frac{}{6} = \frac{}{6}$

$f(1) = {}^4C_1 \cdot \frac{}{2^4} = \frac{4!}{!(4-\ )!} \times \frac{}{6} = \frac{4!}{!3!} \times \frac{}{6}$

Recall 1! = 1

$\frac{4\times3\times2\times}{3\times2\times} \times \frac{}{6} = \frac{}{6}$

$f(1) = \frac{4!}{3!} \times \frac{}{26} = \frac{2}{2!(4!\ 2)!} \times \frac{}{6} = \frac{}{2!2!} \times \frac{}{6} = \frac{4\times3\times2\times}{2\times\ \times2\times} \times \frac{}{6} = \frac{6}{6}$

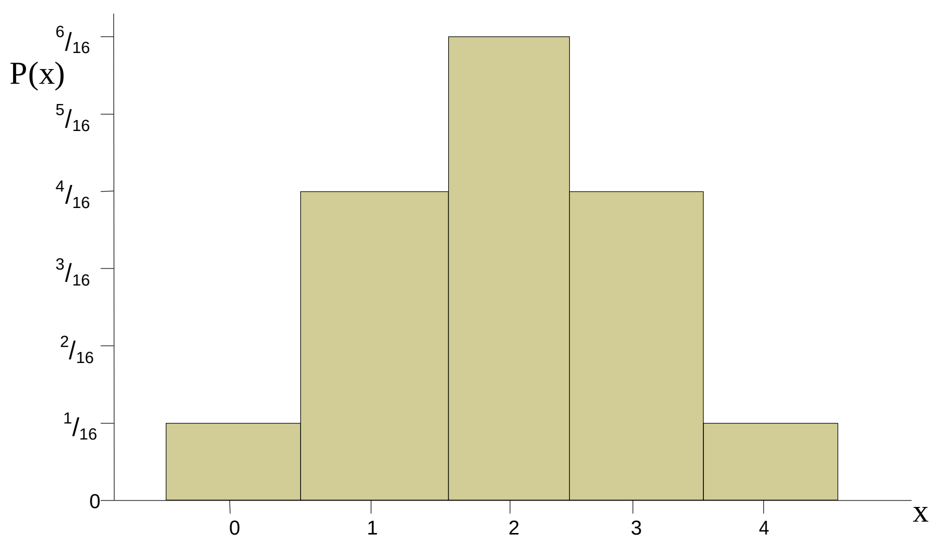
$f(2) = {}^4C_2 \cdot \frac{}{2^4} = \frac{4!}{3!(4-3)!} \times \frac{}{6} = \frac{4!}{3!\ !} \times \frac{}{6} = \frac{4\times3\times2\times}{3\times2\times\ \times\ !} \times \frac{}{6} = \frac{}{6}$

$f(3) = {}^4C_3 \cdot$

$f(4) = {}^4C_4 \cdot \frac{}{2^4} = \frac{4!}{4!(4-4)!} \times \frac{}{6} = \frac{4!}{2!2!} \times \frac{4\times3\times2\times}{6\ 2\times\ \times2\times} \times \frac{}{6} = \frac{}{6}$

b.

X	0	1	2	3	4
P(X)	$^1/_{16}$	$^4/_{16}$	$^6/_{16}$	$^4/_{16}$	$^1/_{16}$



CLASS ASSESSMENT EXERCISE 2

Given that three coins are tossed simultaneously:

(a) Obtain the sample size.
(b) Obtain the probability distribution obtaining *x* where *x* is the number of heads.
(c) Use the value obtained in (b) above to construct the probability histogram for the distribution.

3.3 Random Variable Distribution

The probability distribution of a random variable x provides the possible values of the random variable and their corresponding probabilities. A probability distribution can be in form of a table, a graph or a mathematical formula. The probability distribution of a discrete random variable is a graph or a table or formula that specifies the probability associated with each possible value that the random variable can assume. It can also be described as the listing of all possible values and corresponding probabilities of a discrete random variable or a formula for the probabilities.

The two major requirements for a discrete probability distribution are as follows:-

Let $p(X = x)$ denote the probability the random variable X equals x, then

(i) $0 \leq p(X = x) \leq 1$
(ii) $\sum p(X = x) = 1$

This implies that, each probability takes a value between 0 and 1 and the sum of all the probabilities equals to 1

Examples

A discrete random variable x can assume five possible values: 2, 3, 4, 5, 8, and 10. Its probability distribution is shown as follows

X	2	3	5	8	10
$p(x)$	0.15	0.10		0.25	0.25

(a) What is P(5)?
(b) What is the probability that x equals 2 or 10?
(c) What is $p(x \leq 8)$?

Solution

(a) P(5) = 1 – (0.15 + 0.10 + 0.25 + 0.25)
= 1 – 0.75 = 0.25
Note: P(2) + P(3) + P(5) + P(8) + P(10) = 1
∴ P(5) = 1 – [P(2) + P(3) + P(8) + P(10)]

(b) P(2) or P(8) = 0.15+ 0.25 = 0.40

(c) $P(X \leq 8) = P(2) + P(3) + P(5) + P(8)$
= 0.15 + 0.10 + 0.25 + 0.25 = 0.75

2. Given that f(x) = $\binom{4}{x} \cdot \frac{}{2^4}$. Find

(a) $p(2 \leq x < 4)$

(b) $p(0 < x \leq 3)$

Solution

$$\binom{4}{x} \cdot \frac{}{2^4} = {}^4C_x \cdot \frac{}{6}$$

$$f(0) = {}^4C_0 \cdot \frac{}{6} = \frac{}{6}$$

$$f(1) = {}^4C_1 \cdot \frac{}{6} = \frac{4}{6}$$

$$f(2) = {}^4C_2 \cdot \frac{}{6} = \frac{6}{6}$$

$$f(3) = {}^4C_3 \cdot \frac{}{6} = \frac{4}{6}$$

$$f(4) = {}^4C_4 \cdot \frac{}{6} = \frac{}{6}$$

∴ (a) $p(2 \leq x < 4) = p(2) + p(3)$

$$= \frac{6}{6} + \frac{4}{6} = \frac{0}{6} = \frac{5}{8}$$

(b) $p(0 < x \leq 3) = p(1) + p(2) + p(3)$

$$= \frac{4}{6} + \frac{6}{6} + \frac{4}{6} = \frac{4}{6} = \frac{7}{8}$$

3. Given that f(x) = $\frac{x+2}{25}$. Is it a discrete probability distribution function if the value of x ranges from 1 to 5?

Solution

$$f(x) = \frac{x+2}{25}$$

$$f(1) = \frac{+2}{} = \frac{3}{}$$

$25 \quad 25$

$f(2) = \frac{2+2}{25} = \frac{4}{25}$

$f(3) = \frac{3+2}{25} = \frac{5}{25}$

$f(4) = \frac{4+2}{25} = \frac{6}{25}$

$f(5) = \frac{5+2}{25} = \frac{7}{25}$

Hence, $f(x) \geq 0$ or $0 \leq P(x) \leq 1$ and

$$\sum \mathbf{f}(x) = \frac{3}{25} + \frac{4}{25} + \frac{5}{25} + \frac{6}{25} + \frac{7}{25} = \frac{25}{25} = 1$$

Since the two conditions are satisfied, then, $f(x) = \frac{x+2}{25}$ is a discrete probability function.

4. Find the value of ‘c’ for each of the following to satisfy the condition for being a discrete probability function: -

   (a) $f(x) = C \left(\frac{\;}{2x}\right)^2$; x = 1, 2, 3.

   (b) $f(x) = \frac{x^2}{c}$ $0 \leq x \leq 3$

   (c) $f(x) = C(x+2)^2$ x = 0, 1, 2, 3.

Solution

(a) $f(x) = C \left(\frac{\;}{2x}\right)^2$

$\mathbf{f}(1) = C \left(\frac{\;}{2(\;)}\right)^2 = \frac{\;}{4}C$

$\mathbf{f}(2) = C \left(\frac{\;}{2(2)}\right)^2 = \frac{\;}{6}C$

$\mathbf{f}(3) = C \left(\frac{\;}{2(3)}\right)^2 = \frac{\;}{36}C$

$\therefore \frac{\;}{4}C + \frac{\;}{6}C + \frac{\;}{36}C = 1$, multiply through by 144 (the L.C.M.)

$36C + 9C + 4C = 144$

$49C = 144$

$C = \frac{44}{49} = 2.94$

(b) $f(x) = \frac{x^2}{c}$

$f(0) = \frac{0^2}{c} = \frac{0}{c} = 0$

$f(1) = \frac{\;^2}{c} = \frac{\;}{c}$

$f(2) = \frac{2^2}{c} = \frac{4}{c}$

$f(3) = \frac{3^2}{c} = \frac{9}{c}$

$\therefore \frac{}{c} + \frac{4}{c} + \frac{9}{c} = 1$; multiply through by c,

1 + 4 + 9 = c

c = 14

(c) $f(x) = C(x + 2)^2$

$f(0) = C(0 + 2)^2 = 4C$

$f(1) = C(1 + 2)^2 = 9C$

$f(2) = C(2 + 2)^2 = 16C$

$f(3) = C(3 + 2)^2 = 25C$

$\therefore 4C + 9C + 16C + 25C = 1$

$54C = 1$

$C = \frac{}{54}$

CLASS ASSESSMENT EXERCISE 3

1. The random variable x has the discrete probability distribution shown here:

X	-2	-1	0	1	2
P(x)	0.10	0.15	q	0.3	0.05

(a) Find the value of q.
(b) Find $p(x > -1)$
(c) Find $P(x < 2)$
(d) Find $P(-1 < x < 2)$
(e) Find $P(x < 1)$
(f) Find $P(-1 \leq x \leq 1)$

2. For each of the following, determine the value of constant 'k' so that f(x) satisfy the condition for being a discrete function for a random variable.

(a) $f(x) = k \left(\frac{}{5x}\right)^2$; x = 0, 1, 2.
(b) $f(x) = k(3x - 2)^2$; x $1 \leq x \leq 5$
(c) $f(x) = \frac{k}{\sqrt{x}}$; x = 0, 1, 4, 9, 16
(d) $f(x) = k({}^5C_x)$; x = 0, 1, 2 ... 5
(e) $f(x) = k \left(\frac{}{4}\right)^x$; for x = 0, 1, 2.

3.4 Mean, Standard Deviation and Variance of Random Variables

The mean of discrete random variable is given by the formula:

$\mu_x = \sum[x.\, p(X = x)]$, where x is the value of the random variable and $p(X = x)$ is the probability of observing the random variable x. the mean of a discrete random variable is otherwise known as the expected value of x i.e. $\mu = E(x) = \sum x.\, p(x)$. The expected value is therefore the mean of the probability distribution, or a measure of its central tendency. The mean of a random variable is the long-run average outcome of the experiment. In this sense, as the number of trials increases, the average result of the experiment gets closer to the mean of the random variable.

The standard deviation of a discrete random variable x is denoted by σ_x or σ and it measures the standard deviation of each of the random observation from the mean or the expected value. Therefore

$\sigma = \sqrt{\sum (x - \mu)^2 . p(X = x)}$. The standard deviation of a discrete random variable can also be obtained from the computing formula given as

$\sigma = \sqrt{\sum x^2 . p(X = x) . \mu^2}$ It is therefore called the expected value of the square root of the distance from the mean.

Variance of a random variable x is the square of the standard deviation. It is the expected value of the squared distance from the mean. The variance is often mathematically presented as thus:-

$r^2 = E[(x - Jl)^2] = \sum (x - Jl^2)p(x) = \sum x^2 \ (x)$

Examples

1. Given the ages of 8 students as

   19, 20, 20, 19, 21, 27, 20, 21.

(a) Construct the probability distribution table of X (the ages of a randomly selected student.
(b) Find the mean of the sample.
(c) Find the mean of the discrete random variable
(d) Comment on (b) and (c) above.

Solution

(a)

Ages	**Frequency(f)**	**Probability P(X = x)**
19	2	$\frac{2}{8}$
20	3	$\frac{3}{8}$
21	2	$\frac{2}{8}$
27	1	$\frac{1}{8}$
	$\mathbf{f} = 8$	

$$\overline{\qquad\qquad 8 \qquad\qquad} \qquad \overline{8}$$

(c) $Jl = \sum x\ .p(x)\ = \frac{2}{8} \times 19 + \frac{3}{8} \times 20 + \frac{2}{8} \times 21 + \frac{}{8} \times 27$

$$= \frac{}{8} + \frac{}{8} + \frac{}{8} + \frac{}{8} = \frac{}{8} = 20.875$$

(d) (b) and (c) above are the same i.e. the mean of the sample equals to the mean of the discrete random variable.

2. Given the table below

X	0	1	2	3	4	5	6
P(X = x)	0.029	0.049	0.078	0.155	0.212	0.262	0.215

Find the standard deviation and the variance for the discrete random distribution given that $Jl = 4.118$

Solution

X	P(X = x)	x^2	x^2 P(X = x)
0	0.029	0	0.00
1	0.049	1	0.049
2	0.078	4	0.312
3	0.155	9	1.312
4	0.212	16	3.392
5	0.262	25	6.550
6	0.215	36	7740
			19.438

Standard deviation = r = $\sqrt{\sum(x^2.p(x)\ - Jl^2}$

$$= \sqrt{19.438 - (4.118)^2} = 1.6$$

Variance = $(1.6)^2 = r^2 = 2.56$

3. Given the probability distribution below

X	0	10	20	30
P(x)	1	1	3	1

$\overline{6}$	2	10	30

Find (i) E(x) or the mean

(ii) $E(x^2)$

(iii) $E(x - \bar{x})^2$ i.e. variance

(iv) The standard deviation.

Solution

(i) $E(x) = 0 \times \frac{}{6} + 10 \times \frac{}{2} + 20 \times \frac{3}{0} + 30 \times \frac{}{30}$

$= 0 + 5 + 6 + 1 = 12$

(ii) $E(x^2) = 0^2 \times \frac{}{6} + 10^2 \times \frac{}{2} + 20^2 + \frac{3}{0} + 30^2 \times \frac{}{30}$

$= 0 + 50 + 120 + 30 = 200$

(iii) $E(x - \bar{x})^2$ = variance

$= [(0 - 12)^2 \times \frac{}{6} + [(10 - 12)^2 \times \frac{}{2}] + [(20 - 12)^2 \times \frac{3}{0}] + [(30 - 12)^2 \times \frac{}{30}]$

= 24 + 2 + 19.2 + 10.8 = 56 = variance

(iv) Standard deviation = $\sqrt{\text{i1ft1 CR}} = \sqrt{56} = 7.48$

CLASS ASSESSMENT EXERCISE 4

1. Given the table below:

Y	0	1	2	3	4	5
P(Y = y)	0.424	0.161	0.134	0.111	0.093	0.077

Find the mean standard deviation and he variance for the discrete random distribution.

2. A bag contains 3 red balls and 5 white balls. If two balls are drawn without replacement;
   (a) Construct a probability distribution of a random variable *x* where *x* is the number of red balls drawn.
   (b) Compute the mean and the standard deviation for the discrete random distribution.

3.5 Mathematical Expectation

Mathematical expectation arises in connection with games of chance and in its simplest form, it is the product of the amount a player stands to win and the probability attached to winning. Mathematical expectation of a random variable is simply its expected value. If *x* is a discrete random variable and f(x) is the value of its probability distribution at *x*, the expected value of this random variable is

$E(x) = \sum x \cdot f(x)$. For instance, if a person invests some amount (s), in a viable business and the probability of his making a profit is p, then mathematical expectation is p . s (product of *p* and *s*).

Example

1. The probability of many receiving a gift of ₦ 500 from her uncle is $^3/_5$. Find the mathematical expectation.

Solution

$P = {^3/_5}$, S = ₦500

$\therefore$ Mathematic expectation = p . s = ₦500 × $^3/_5$ = ₦ 300

2. What is the fair price (expectation) to enter a game in which one can win ₦3000 with a probability 0.25and also ₦500 with a probability of 0.1.

Solution

P = (₦3000 × 0.25) + (₦500 × 0.1)

= ₦750 + ₦500 = ₦1,200

3. A trader can make a profit of ₦2000 with a probability of 0.7 and a loss of ₦500 with a probability of 0.5. Determine her expectation.

Solution

(₦2000 × 0.7) – (₦500 × 0.5)

= ₦1400 – ₦250 = ₦1,150

CLASS ASSESSMENT EXERCISE 5

1. If the probability that a bag will receive ₦100,000 from his daddy is ¾. Find his mathematical expectation?
2. If it rains, an umbrella salesman earns ₦200 per day as commission and if it does not rain, he loses ₦50 per day. If the chance that it rains is $^2/_3$, calculate the expectation of salesman.

4.0 CONCLUSION

This unit has been able to examine the meaning and scope of random variables and its distribution. The mean, standard deviation and variance of a discrete random variable has also been treated. The concept of mathematical expectation with respect to discrete probability concludes the unit.

5.0 SUMMARY

Random variables describe a numerical measure of the outcome of a probability experiment whose value(s) is determined by chance. Such variable could be discrete or continuous. The mean, standard deviation and the variance of discrete and continuous distribution can be obtained. The concept of mean of discrete random distribution is alternatively called Expected Value and it is particularly useful in obtaining the mathematical expectation.

6.0 TUTOR MARKED ASSIGNMENT

1. Specify whether each of the following random variable is continuous or discrete.

   (a) The number of words spelt correctly by a student in a spelling test.

   (b) The amount of water flowing through a Dam in a day.

   (c) The length of time an employee is late for work

   (d) Your weight

   (e) The amount of carbon monoxide produced per gallon of unleaded gas.

2. Given the distribution below: -

X	0	1	2	3	4	5
P(x)	0.002	0.029	0.132	0.309	0.360	0.168

   (a) Find $\mu = E(x)$. Interpret the result.
   (b) Find $\sigma = \sqrt{E(X - \mu)^2}$. Interpret the result.
   (c) Find the variance.

3. Given that y is the sum of scores obtained from tossing two dices simultaneously.
   (a) Complete the table below

Y	2	3	4	5	6	7	8	9	10	11	12
P(Y = y)											

   (b) Find the E(x), σ and σ^2.
   (c) Draw the probability histogram for the distribution

4. Construct a probability histogram for each of the following discrete probability distribution.

   (a) $f(x) = \frac{\binom{2}{x}\binom{4}{3-x}}{\binom{6}{3}}$ where x = 0, 1, 2.

   (b) $f(x) = {}^5C_x \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{5-x}$ for $0 \le x \le 5$

5. For each of the following, determine the value of c so that f(x) satisfies the condition for being a discrete random variable.
   (a) f(x) = $\frac{2}{5x}$ C; x = 0, 1, 2
   (b) f(x) = $C(2^{x})$ **fof** $1 \leq x < 4$
   (c) f(x) = $\frac{c}{\sqrt[3]{x}}$ **fof** $x = 1, 8, 27, 64$

7.0 REFERENCES/FURTHER READINGS

Loto, Marget A. et al (2008). Statistics Made Easy. Concept Publication Limited.

Sincich, McClaire and McClave, J. T. (2009). Statistics. Pearson International Edition. (11th Edition). Pearson Prentice Hall, Saddle River, USA.

Sullivan, Micjeal (2005). Fundamentals of Statistics. Pearson Prentice Hall, Upper Saddle River, New Jersey, USA.

Weiss, Neil A. (2008). Introductory Statistics. Pearson International Edition. (Eighth Edition). Montreal Munich Publishers, Singapore.

Module 6: Index Number

Unit 1: Meaning and Scope of Price Index Number

Unit 2: Weighted Index Numbers

Unit 3: Other Index Numbers

UNIT 1

MEANING AND SCOPE OF PRICE INDEX

1.0 INTRODUCTION

Economic variables changes over time and space. For example, prices of commodities may increase at one time, remain constant or decrease at another time. So many changes occur in industrial output, money in circulation, foreign trade, agricultural output, etc. A device to measure the changes that take place in the individual economic variables is called the index number.

This unit shall introduce you to the meaning, scope and uses of price number. The concept of price relative index number with respect to its various types shall also be discussed in this unit.

2.0 OBJECTIVES

At the end of the unit, you should be able to:-

(i) Explain the meaning and uses of price numbers; and
(ii) Identify and compute the various properties of price relative price index number.

3.0 MAIN CONTENT

3.1 Meaning and uses of Price Number

An index number can be defined as statistical measures of a variable or a group of related variables with respect to its value at a specified period. This specified period is known as the base year. It is a means for estimating trends in price, wages, production and other related economic variables over a period of time.

At one time or the other, either in the mass media or in everyday discussion, we have come across terms such as “consumer price index”, “unemployment index”, “production index”, etc. These terminologies are often used in business and economic situation. They all find their origin from the concept of a statistical measure known as the index number.

When comparing series of figures, it is often likely that some complexity arises in making direct comparison. To take care of this embarrassment, what we need is a single figure which in itself will show how much one year differs from another. A convenient way of doing it is to take a fairly typical year as a base, and expresses the figures for other year as a percentage of this. For example, if the figure for year 2000 was 100 and that for 2001, 108, we should know that production (or whatever) was 8% higher. Such a single figure summarizing a comparison between two sets of figure is called an index number.

It is a weighted average compiled from a selection of item representing the relative variable. For example, it may be an average of the changes in the price of the individual items in the group. In the case of price index or in the case of quantity index, it may be an average of the changes in the quantity of the individual items in the group. However, an index number must be a single figure that represents the net effect of the change in the constituent variables.

Uses of Index Number

The use of index number in business and economic situations is quite rewarding. Apart from providing general information, it enables one to forecast business and economic conditions. It also enhances comparison of related items from one year to the other or from one part of the world to another. Index numbers are also used to measure changes in the prices of commodities, changes in volume of industrial output, fluctuations in foreign trade activities, etc. specifically, index numbers are used:-

(i) To show the trend which enable comparisons to be made in respect of inter and intra commodity price movement;
(ii) To measures changes between the level of prices and wages;
(iii) To measures changes between the levels of production and wages;
(iv) To measures the changes between the level of domestic production and imports;
(v) To measures the changes between the level of imports and exports;
(vi) To show trend which may enable comparison to be made over space between birth rates, death rates, etc

CLASS ASSESSMENT EXERCISE

1. Explain the concept of Index Number.
2. Outline the usefulness of Index Number in economic and business.

3.1 Price Relative Price Index Number

This type of index number is one of the simplest types. It involves dividing the price of a given term in one year (P_n) by its price in a specified year known as the base year (P_o) and expressing this ratio as a percentage. So, price relative = $\frac{p}{p_o}$,

P_n = current or given year price for the item;

P_o = base year price for the item.

If x is the base year and y is the given year, then we sometimes write the price relative for y given x as the base year as follows:-

$$p_x = \frac{p}{p_x}$$

Example 1

The price of a television set in 1996 and 1997 were N46,000 and N62,000 respectively.

(a) Compute the price relative using 1996 as the base year.
(b) Compute the price relative using 1997 as the base year.

Solution

Price in 1996 = N46, 000

Price in 1997 = N62, 000

(a) Price relative = $p_{1996 \quad 1997} = \frac{\text{pftCR t} \quad 1997}{\text{pftCR t} \quad 1996} \times \frac{100}{1}$

$$= \frac{62000}{46000} \times \frac{100}{1}$$

$$=134.8\%$$

This implies that there has been an increase of (134.8 - 100) 34.8% over the price in 1997.

(b) If 1997 is used as a base year, the price relative is

$$p_{1996 \quad 1997} = \frac{\text{pftCR t} \quad 1996}{\text{pftCR t} \quad 1997} \times \frac{100}{1}$$

$$= \frac{46000}{62000} \times \frac{00}{}$$

$$= 74.2\%$$

This means that the price in 1996 was 74.2% of that of 1997 or there is a decrease or reduction of (100 – 72.4) 25.8% of the price in 1996 when compared with that of 1997.

Properties of Price Relative

Price relative satisfy the following properties:

(i) **Identity property: -** The price relative for a given period with respect to the same period is 100% or 1. It is expressed in percentage $p_{a/a} = 1 = 100\%$

(ii) **Time Reversal Property:** - If two periods are interchanged, then the corresponding price (if not expressed in percentage) are reciprocal of each other, as $p_{a/b} = {}^{1}/p_{b/a}$.

In the example above, the price relative for 1997 given 1996 as base year (if not expressed in percentage) is $\frac{62000}{46000} = \frac{46}{62}$, while the relative for 1996 given 1997 as base year (and not expressed on percentage) is $\frac{46000}{62000} = \frac{62}{46}$, which is the reciprocal of the value obtained if 1996 is the base year. So, $p_x = \frac{1}{p\ \ x}$.

(iii) **Cyclical Properties:** - If x, y and z are three given periods, then $p_x \,.\, p_z \,.\, p_{z/x} = 1$

Note that $p_x \,.\, p_z$ is a chain relative.

$$p_x \,.\, p_z \,.\, p_{z/x} = \left(\frac{\text{pftCR t} \quad \text{Rfto} \quad y}{\text{pftCR t} \quad \text{Rfto} \quad x}\right) \times \left(\frac{\text{pftCR t} \quad \text{Rfto} \quad z}{\text{pftCR t} \quad \text{Rfto}}\right) \times \left(\frac{\text{pftCR t} \quad \text{Rfto} \quad x}{\text{pftCR t} \quad \text{Rfto} \quad z}\right)$$

$$= \frac{P_y}{P_X} \times \frac{P_z}{P_y} \times \frac{P_X}{P_z} = 1$$

The third property can be extended to more than three periods and can also be modified to obtain a modified cyclical property.

The popularly used average is the arithmetic mean, although the median or geometric mean of the price relatives of the item in the group may be used. If the arithmetic mean is used, the simple average of relative price index is $\frac{1}{N} \cdot \sum \frac{P_n}{P_0}$

Example 2

The following is the prices of tins of Groundnut oil, Soya beans oil, and Palm Oil for 1990 to 1993.

	1990	**1991**	**1992**	**1993**
Groundnut oil	600	680	700	900
Palm oil	500	510	520	600
Soya bean oil	700	720	800	1000

Use the simple average of relative method to compute a price index of oil in 1993 using:

(a) 1990 as base year

(b) 1990-1992 as base year

Solution

(a) Using 1990 as base year,

For 1993: Price relative of groundnut oil

For 1993; $\frac{900}{600} \times \frac{100}{1} = 150$

Price relative of palm oil

For 1993; $\frac{600}{500} \times \frac{100}{1} = 120$

Price relative of soya bean

For 1993; $\frac{1000}{700} \times \frac{100}{1} = 142.9$

$\therefore$ Average of price relative = $\frac{150 + 120 + 142.9}{3}$

= 137.6

(b) If 1990-1992 is to be used as base year, we use to obtain a value for the base for each item.

Average (1990-1992): Groundnut = $\frac{600 + 6800 + 700}{3} = 660$

(i) $\therefore$ Relative price for Groundnut in 1993 using1990-1992 as base year

$= \frac{900}{660} \times \frac{100}{1} = 136.4$

Average (1990-1992): Palm oil = $\frac{500 + 510 + 520}{3} = 510$

(ii) $\therefore$ Relative price for Palm oil in 1993 using1990-1992 as base year

$= \frac{600}{510} \times \frac{100}{1} = 117.6$

Average (1990-1992): Soya bean oil = $\frac{700 + 720 + 800}{3} = 740$

(iii) $\therefore$ Relative price for Soya bean oil in 1993 using1990-1992 as base year

$= \frac{1000}{740} \times \frac{100}{1} = 135.$

$\therefore$ Price index = average of (i), (ii) and (iii)

$= \frac{136.4 + 117.6 + 135.1}{3} = \frac{139.1}{3} = 129.7$

Note: - The median value could also be used instead of the average (mean). If this is done, in question (a), the value obtained is 142.9 while the final answer obtained in (b) is 135.1

Example 3

Given the table below:

Commodity	**Average Price per kg**	
	1980	**1985**
Beef (kg)	₦ 4.50	₦ 6.00
Fish (kg)	₦ 2.80	₦ 3.35
Chicks(kg)	₦ 390	₦ 4.75

From the table, calculate the simple aggregate price index.

Solution

Simple aggregate price index = $\frac{\Sigma}{\Sigma_{o}} \times \frac{100}{1}$

$$= \left(\frac{6.00 + 3.35 + 4.75}{4.50 + 2.80 + 3.90}\right) \times \frac{100}{1}$$

$$= \frac{14.10}{11.20} \times \frac{100}{1} = 125.9$$

CLASS ASSESSMENT EXERCISE 2

Given the table below: -

Commodities	**Year (Prices (₦))**			
	1970	**1980**	**1990**	**2000**
Yam (kg)	15	18	20	22.50
Fish (kg)	20	22.50	25	26.50
Meat (kg)	50	60	75	80
Eggs (crates)	40	30	25	15

(a) Compute the price relative for each commodity in 2000 using 1980 as the base year. Interpret your answer.

(b) Compute the simple aggregate price index in 1990 using 1970 as the base year.

(c) Use the simple average of relative method to compute the price index of the commodities (aggregate) in 2000 using;
   (a) 1980 as the base year
   (b) 1970-1990 as the base year.

4.0 CONCLUSION

This unit has been able to introduce the students to the meaning and concept of Price Number. An attempt has also been made to expose students to the price relative index number (the simple form of index number). Various properties of price relatives have been discussed. Attempts have also been made at exposing the students to simple mathematical problems relating to price index and the interpretation of the result.

5.0 SUMMARY

Index Number is a statistical measure of variable or a group of relative variables, with respect to its value at a specified period of time known as the base year. The concept is useful in economics, business and finance in the prices of making comparison and relative judgement between/among the variables concerned with respect to changes in time in the values or variable concerned.

The simplest form of index number is the price relative which shows the changes in price of goods and services over a given period of time. This concept is guided by a number of properties, which include identity property, time reversal property, cyclical property etc.

6.0 TUTOR MARKED ASSIGMENT

1. The table below shows the price of a commodity for a period of 5 years

Year	1985	1986	1987	1988	1989
Price (₦)	48	63	57	75	90

(i) Using 1985 as a base year, find the price relative corresponding to 1987 and 1989.
(ii) Using 1986 as a base year, find the price relative corresponding to each of the given years.
(iii) Using 1987-1989 as a base year, find the price relative corresponding to each of the years.

2. Given the table below, compute the chain index number (a – f).

Year	**Price (₦)**	Chain Index
1999	12,912	a b c d e f
2000	18,671	
2001	21,200	
2002	28,633	
2003	35,028	
2004	40,650	
2005	44,531	

Hint: - A chain index could simply be defined as an ordinary price index in which each period in the series use the previous period as base year, e.g. a $=\frac{18671}{12912} \times \frac{100}{1} = 145$

7.0 REFERENCES / FURTHER READING

Loto, Margeret A. et al (2008): Statistics Made Easy. Concept Publication Limited, Lagos.

UNIT 2

WEIGHTED INDEX NUMBER

Table of Contents

1.0 INTRODUCTION

In the last unit, simple aggregate method and simple average relative method of price changes are considered. They do not take into consideration the relative importance of the various items since they both assume equal weight for the items. To overcome this inadequacy, the weighted price index and the quantity of volume relative were designed. This unit highlights the various approaches or methods involved in obtaining quantity relative and weighted price index numbers.

2.0 OBJECTIVES

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

(i) Compute and interpret quantity relatives

(ii) Compute and interpret weighted price index numbers using various approaches or methods.

3.0 MAIN CONTENT

3.1 Volume Index

When the quantities or volume of commodities are to be compared instead of the price of commodities, then we talk about quantity or volume relatives. Examples are volume of import, volume of production of cement, volume of production of oil, etc. Unlike the price indices, here, the quantities instead of the prices are used as weights. We have different types of methods of computations. These include:-

(i) Quantity Relative or Volume Relative or Simple Quantity Relative $= \frac{e}{e_o} \times \frac{100}{1}$

(ii) Simple Quantity Index for basket of commodities or Simple aggregate volume or aggregate quantity relative; given as $\frac{\sum e}{\sum e_o} \times \frac{100}{1}$

(iii) Laspeyres' Weighted Quantity or Volume Index, given as

$$\frac{\sum e \quad _o}{\sum e_o \quad _o} \times \frac{100}{1} = L_W$$

(iv) The Paasche's weighted quantity volume index number, given as:

$$\frac{\sum e}{\sum e_o \quad _o} \times \frac{100}{1} = p_w$$

(v) The Fisher's ideal quantity index = $\mathbf{j}\,\overline{L_W . p_W}$ = $\sqrt{\frac{\sum q_n p_o}{\sum q_o p_o} \cdot \frac{\sum p_n q_n}{\sum p_o q_n}} \times \frac{00}{}$

Example

The following are prices (in kobo per litre) and quantities (in million of barrels) of petrol products produces in a certain state of the country in two years: (each barrel contains 159 litres)

	1990		1990	
Products	**Price**	**Quantity**	**Price**	**Quantity**
Kerosine	2.20	22	8	50
Petrol	6.10	15	11	35
Diesel oil	5.40	10	14	25

Calculate:

(a) Simple quantity relative for Petrol;
(b) Simple quantity index for all the products or aggregate volume index;
(c) The Laspeyre's weighted quantity or volume;
(d) The Paasche's weighted quantity' and
(e) The Fisher's volume index; using 1980 as the base year.

Solution

(a) Simple Quantity Relative for Petrol = $\frac{e_{1990}}{e_{1980}} \times \frac{100}{1} = \frac{35}{15} \times \frac{100}{1}$

$= 233.3$

(b) Simple quantity index for all the product = $\frac{\sum e}{\sum e_o}$

$= \frac{50+35+25}{22+15+10} \times \frac{100}{1} = \frac{110}{47} \times \frac{100}{1}$

$= 234$

(c) **Note:** P_o = 1980 price
P_n = 1990 price

q_o = 1980 quantity
q_n = 1990 quantity

(d) ∴ Laspeyre's volume index for 1990 is $\frac{\sum q_n p_o}{\sum q_o p_o} \times \frac{00}{}$

$$= \frac{50 \times 2.20 + 35 \times 6.1 + 25 \times 5.4}{2.2 \times 22 + 6.1 \times 15 + 5.4 \times 10} \times \frac{100}{1}$$

$$= \frac{110 + 213.5 + 135}{48.4 + 91.5 + 54} \times \frac{100}{1}$$

$$L_w = \frac{458.5}{193.9} \times \frac{100}{1} = 236.5$$

(e) The Paasche's volume index = $\frac{\sum e}{\sum e_o} \times \frac{100}{1}$

$$p_w = \frac{50 \times 8 + 11 \times 35 + 14 \times 25}{22 \times 8 + 15 \times 11 + 10 \times 14} \times \frac{100}{1}$$

$$= \frac{400 + 385 + 350}{176 + 165 + 140} \times \frac{100}{1}$$

$$p_w = \frac{1135}{481} \times \frac{100}{1} = 236.0$$

(f) Fisher's ideal volume index = $.\mathrm{J}\overline{L_w . p_w} \times \frac{100}{1}$

$$= \sqrt{236.5 \times 236} \times \frac{100}{1} = 236.2$$

CLASS ASSESSMENT EXERCISE

Given the table below:

Some Food Items	July 1981		July 1983	
	Price (₦)	**Qty (kg)**	**Price (₦)**	**Qty (kg)**
Yam	21.8	2	3.25	7
Cassava	1.80	17	2.10	15
Plantain	1.01	6	2.07	8
Rice	1.09	5	2.15	11
Wheat	0.89	4	1.20	6

From the information in the above table, compute for July 1983 base upon July1981 figure:-

(i) Laspeyre's volume index
(ii) Paasche's volume index
(iii) Fisher's ideal volume index

3.2 Weighted Price Index Number

The weighted price index number is similar in some respect to the volume index except that in the weighted price index, quantity is held constant while changes in price is measured using the quantity as the weight attached. This method uses either the quantity of the base year as the weights or the quantity of the current year as the weight. Three major categories of weighted Price Index Numbers are:-

(i) The Laspayre's which is given by:

$$L \ = \frac{\sum p \ e_o}{\sum \ {}_o e_o} \times \frac{100}{1}$$, where the base year quantity is used as weight.

(ii) The Paasche's price index given by

$$P_p = \frac{\sum e}{\sum \ {}_o e} \times \frac{100}{1}$$, where the quantity of the current year is used as the weight.

(iii) Fisher's Ideal Price Index

This is a compromise price index to the Laspeyre's and Paasche's index numbers. It is defined as the geometric mean of the Laspeyre's and Paasche's index numbers as given

$$F = \sqrt{\frac{\sum p_n q_o}{\sum p_o q_o} \quad \frac{\sum q_n p_n}{\sum p_o q_n}} \times \frac{00}{}$$

It should be noted that while Laspeyre's tends to overestimate price changes, Paasche's tend to understimate the changes. The Fisher's price index takes the value between Laspeyre's and Paasche's and thus a better estimate of price changes than the first two index numbers (Paasche's and Laspeyres'). The Fisher's price index number satisfies the factor reversal test.

Example

The table below shows the price and quantity of oil bought at a black market.

	1992		**1993**		**1994**	
	Price	**Quantity**	**Price**	**Quantity**	**Price**	Quantity
Diesel	₦ 35	100	₦ 40	90	₦ 52	130
Petrol	₦ 40	120	₦ 55	85	₦ 60	100
Engine oil	₦ 55	130	₦ 140	95	₦ 200	80

Calculate the weighted price index using:

(a) The Laspeyre's price index;
(b) The Paasche's price index; and
(c) The Fisher's Ideal Price Index for 1993 and 1994 using 19992 as the base year.

Solution

(a)(i) The Laspeyre's price index using 1992 as base year and 1993 as current year gives:-

$$= \frac{\sum p_n q_0}{\sum p_0 q_n} \times \frac{100}{1} = \frac{(40 \times 100) + (55 \times 120) + (140 \times 130)}{(35 \times 100) + (40 \times 120) + (55 \times 130)} \times \frac{100}{1}$$

$$= \frac{4000 + 6600 + 18200}{3500 + 4800 + 7150} \times \frac{100}{1}$$

$$= \frac{2880}{15450} \times \frac{100}{1} = 186.4$$

(ii) The Laspeyre's price index using 1992 as base year and 1994 as current year gives:-

$$= \frac{\sum p_n q_o}{\sum p_o q_n} \times \frac{100}{1} = \frac{(52 \times 100) + (60 \times 120) + (200 \times 130)}{(100 \times 35) + (40 \times 120) + (55 \times 130)} \times \frac{100}{1}$$

$$= \frac{5200 + 7200 + 26000}{3500 + 4800 + 7150} \times \frac{100}{1}$$

$$= 248.5$$

(b)(i) The Pasche's price index using 1992 as base year and 1993 as current year gives:

$$= \frac{\sum p_n q_n}{\sum p_0 q_n} \times \frac{100}{1} = \frac{(40 \times 90) + (55 \times 85) + (140 \times 95)}{(35 \times 90) + (40 \times 85) + (55 \times 95)} \times \frac{100}{1}$$

$$= \frac{3600 + 4675 + 13300}{3150 + 3400 + 5225} \times \frac{100}{1}$$

$$= \frac{21575}{11775} \times \frac{100}{1} = 183.2$$

(ii)The Paasche's price index using 1992 as the base year and 1994 as the current year is given as

$$= \frac{\sum p_n q_n}{\sum p_0 q_n} \times \frac{100}{1} = \frac{(52 \times 130) + (60 \times 100) + (200 \times 80)}{(35 \times 130) + (40 \times 100) + (55 \times 80)} \times \frac{100}{1}$$

$$= \frac{6760 + 6000 + 16000}{4550 + 4000 + 4400} \times \frac{100}{1}$$

$$= \frac{28760}{12950} \times \frac{100}{1} = 222.1$$

(c)(i)Fisher's ideal price index for 1993 using 1992 as the base year

$$= \sqrt{\left(\frac{\sum p_n q_o}{\sum p_o q_n}\right)\left(\frac{\sum p_n q_n}{\sum p_0 q_n}\right)} \times \frac{100}{1}$$

$$= \left(\sqrt{\frac{2880}{15450} \times \frac{21575}{11775}}\right) \times \frac{100}{1}$$

$$= 1.848 \times 100 = 184.8$$

(ii) The Fisher's ideal price index for 1994 using 1992 as the base year

$$= \sqrt{\left(\frac{\sum p_n q_o}{\sum p_o q_n}\right)\left(\frac{\sum p_n q_n}{\sum p_0 q_n}\right)} \times \frac{100}{1}$$

$$= \left(\sqrt{\frac{38400}{15450} \times \frac{28760}{12950}}\right) \times \frac{100}{1}$$

$$= 2.349 \times 100 = 234.9$$

CLASS ASSESSMENT EXERCISE 2

Given the table below:-

	January 1975		**January 1980**	
	Unit price (₦)	**Quantity**	**Unit price (₦)**	**Quantity**
Bread (loaf)	15	84	22	112
Milk (tin)	11	30	18	43
Meat (kg)	215	120	305	94
Palm oil (litres)	128	45	235	38
Fish (kg)	95	60	120	65
Garri (kg)	42	39	65	56
Rice (kg)	48	58	87	75
Beans (kg)	137	102	270	80

Compute for January 1980 base on January 1975:-

(i) Base weighting (Laspeyre) price index
(ii) Current weighting (Paasche) price index
(iii) The Fisher's ideal price index

4.0 CONCLUSION

This unit considers index numbers with respect to volume of commodities as well as weights attached in terms of quantity. The weight index has been discussed in terms of attaching the weights either to the price or to the quantity. In both cases, various dimensions (Laspeyre's, Paasche's and the Fisher's ideal indices) has been treated.

5.0 SUMMARY

Index numbers are not limited to price relatives alone. Indices relating to quantity changes can also be obtained with the use of volume or quantity relative, simple quantity index, Laspeyre's weighted quantity, Paasche's weight quantity and the Fisher's ideal quantity index.

Weights attached in terms of quantity to prices are essential for effective and efficient comparison about the changes in the volume of sales or turn over. The weighted price index number addresses the shortcoming of price relative by considering the weight (quantity) attached to alternative prices. In this approach, the quantity of the base year may serve the purpose of the weight (Laspeyre's weighted price index) or the quantity of the current year may serve the purpose (Paasche's price index). In order to minimize the shortcoming of Laspeyre's and Paasche's price indices, the geometric mean of the two may be used as contained in the Fisher's ideal price index.

6.0 TUTOR MARKED ASSIGNMENT

1. Given the information below:-

Products	**1995**		**1996**		**1997**	
	Price ($)	**Quantity**	**Price $**	**Quantity**	**Price $**	**Quantity**
Gold (kg)	56	100	65	150	75	200
Mercury (kg)	60	120	95	100	100	230
Diamond (kg)	120	160	190	120	250	280

Required:

(a) Compute the weighted price index using
  (i) Laspeyre's approach
  (ii) Paasche's approach
  (iii) Fisher's ideal index
(b) Compute the;
  (i) Laspeyre's volume index
  (ii) Paasche's volume index
  (iii) Fisher's volume index

2. A large manufacturer purchases an identical component from three independent suppliers that differ in unit price and the quantity supplied. The relevant data for 1993 and 1995 is given as follows

Supplier	**Quantity (1993)**	**Unit price**	
		1993	**1995**
A	150	5.45	6.00
B	200	5.60	5.95
C	120	5,50	6.20

Compute the weight aggregate index using 1993 as the base year.

7.0 REFERENCES / FURTHER READING

Loto, Margeret A. et al (2008): Statistics Made Easy. Concept Publication Limited, Lagos.

UNIT 3

OTHER INDEX NUMBER

Table of Contents

1.0 INTRODUCTION

In the first and the second units of this module, you have been introduced to price relative index number, volume index and the weighted price index. These are the basic index number. Other forms of index numbers that are of equally relevant to economics, business and finance are the Value index and the Consumer price index.

This unit shall conclude the module by discussing these other components of index numbers.

2.0 OBJECTIVES

At the end of the unit, you should be able to:-

(i) Estimate value index given a set of information;
(ii) Estimate consumer price index or retailer price index and use it to interpret the rate of inflation;
(iii) Calculate physical volume index; and
(iv) Outline the problem of index number.

3.0 MAIN CONTENT

3.1 Value Index

This type of index number takes care of both changes in price and quantity and can thus be defined as a statistical measure of change of standard of living over a period of time. Suppose p_o and q_o are the price and quantity of a commodity during a given base year and p_n and q_n denote the current or given price and quantity respectively, then;

Value relative = $\frac{v_n}{v_o} = \frac{p_n q_n}{p_o q_o} = \left[\frac{p_n}{p_o}\right]\left[\frac{q_n}{q_o}\right].$

The value index is easily recognized as the product of the price relative index and the quantity relative index. It has the properties of price and quantity relatives. In addition, it satisfies the factor intensity reversal. When expressed in percentage, it is given as;

$$\frac{\sum p_n q_n}{\sum p_o q_o} \times \frac{100}{1}$$

Example

Given the above information

Products	1980		1990	
	Price (₦)	**Quantity (Litres)**	**Price (₦)**	**Quantity (Litres)**
X	2.2	22	8	50
Y	6.10	15	11	35
Z	5.40	10	14	25

Require: Obtain the value index in 1990 using 1980 as the base year.

Solution

Value index = $\frac{\sum p_n q_n}{\sum p_o q_o} \times \frac{100}{1}$

$$= \frac{(50 \times 8) + (35 \times 11) + (14 \times 25)}{(22 \times 2.2) + (6.10 \times 15) + (5.4 \times 10)} \times \frac{100}{1}$$

$$= \frac{1135}{193.9} \times \frac{100}{1}$$

= 585.3

CLASS ASSESSMENT EXERCISE 1

The following is the price of products A, B, and C for 1990 to 1993.

Products	1990		1991		1992		1993	
	Price(₦)	**Qty**	**Price(₦)**	**Qty**	**Price(₦)**	**Qty**	**Price(₦)**	**Qty**
A	5	100	7	160	12	150	6	10
B	8	110	8	80	15	150	9	20
C	10	115	10	70	20	140	5	30

Required: -

Obtain the value index.

3.2 Consumer Price Index and Rate of Inflation

The consumer price index number or the cost of living index numbers are the indices dealing with changes in prices of goods and services need by the average person. Example of these goods and services are housing, electricity consumption, food, drinks medical care, education, social benefits etc. There are different types of consumer price index (CPI), among are CPI for urban areas, CPI for rural area, composite CPI, etc. The Bureau of Statistics in Nigeria collects the data and computes various CPIs. The procedure is to collect results from samples of the focus of interest and used technique of probability sampling to obtain the data required for computation of CPI, which is actually done using Laspayre's Price Index Number

Consumer price index helps in the computation of inflation rate. Sometimes, one's income may be higher than that of previous years, but instead of the purchasing power to increase, it decreases due to inflation. The apparent high salary is not the real income. Rather, it is known as ***apparent or physical or nominal income***. The real income can be computed by dividing the nominal or apparent income by the CPI, using the appropriate base year period, that is $\frac{Current \quad Income}{Current \quad CPI} \times \frac{100}{1}$

Inflation rate can also be computed from the Consumer Price Index. It (inflation) is given as: - $\frac{CPI(Current \quad Year) - CPI(BaseYear)}{CPI \quad (BaseYear)} \times \frac{100}{1}$

Example

A man earned N15, 000 in 1990 and N25, 000 in 1995. If the average CPI (1990 = 100) for 1995 was 320;

(a) What is the physical salary in 1995?

(b) What is his real income in 1995?

(c) What is the normal increase in income in 1995?

(d) What is the real increase in income in 1995?

(e) How much would the worker have earned in 1990, if CPI = 320 in 1995 using 1990 as the base year?

Solution

(a) Apparent/Physical salary in 1995 = ₦ 25,000

(b) His real income = $\frac{\text{Apparent Income}}{\text{CPI}} \times \frac{00}{}$

$$= \frac{25000}{320} \times \frac{00}{} = ₦7,812.50$$

(c) The nominal increase in income = ₦ 25, 000 – ₦15, 000

= ₦ 10, 000

(d) The real increase in 1995 = ₦ 7,812.5 – ₦ 15, 000

= – ₦ 7187.50

This implies that instead of an increase we have a decrease in the real salary in 1995 compared with the average of 1990.

(e) For the worker to receive real salary of ₦ 15,000 in 1995 despite CPI of 320, then

$$\frac{\text{CqffR a CpI}}{\text{b1sR CpI}} = \frac{\text{CqffR a 1tf1}}{\text{b1sR R1f 1tf1}}$$

$$\frac{320}{00} = \frac{\text{current naira}}{5{,}000}$$

$$\frac{320}{00} = \frac{x}{5{,}000},$$

cross multiply, 100 = 320 x 15,000

= ₦48, 000.

So, he needs to earn ₦48, 000 in 1995 in order to be at par with the 1990 income.

Example 2

Below is a table of Retail Price Index (Average 1982)

Items	**Indices Jan. 1980 = 100**	**Weight**	
		Pensioners	**All household**
Food	138.2	310	290
Drink	131.8	50	85
Tobacco	139.8	70	83
Clothing	117.1	140	80

(a) From the table, compute separate Retail price index for the pensioners and all household.

(b) Why the difference in the values obtained in (a) above.

Items	**Index deviation (D) 19980 = 100**	**Pensioners**		**All household**	
		Weight (w)	**WD**	**Weight (w)**	**WD**
Food	138.2-100=38.2	310	11842	290	11078
Drink	131.8-100=31,8	50	1590	85	2703
Tobacco	139.3-00=39.3	70	2751	83	3261.9
Clothing	117.1-100=17.1	140	2394	80	1368
		570	18577	538	18410.9

Adjustment (Pensioners) = $\frac{\sum WD}{\sum W} = \frac{8577}{570} = 32.59$

Adjustment (All Household) = $\frac{\sum WD}{\sum W} = \frac{84 \quad 0.9}{538} = 34.22$

Retail Price Index (Pensioner) = 100 + 32.59 = 132.59

Retail Price Index (All household) = 100 + 34.22 = 134.22

(b) The two indices differ due to the differences in the weighting, reflecting pattern of expenditure.

CLASS ASSESSMENT EXERCISE

1. The expenditure of a certain family in 1992 with 1991 as base are as follow

Item	Price Relative	Weight
Housing	a	2
Medical care	50	1
Food	b	3
Others	150	4
Total	600	

If the cost of living index of the family in 1991 was 167, find the values of 'a' and 'b'.

2. If the CPI for 1995 and 1996 are 320 and 350 respectively. Find the inflation rate n 1996.

3.3 Problems of Computing Index Numbers

When computing index number, the use of a base year is very important. It is desirable that the base is such that the price of items in the basket be stable and the quantities reasonable. In practice such a perfect base year is impossible. This implies that no index number is free from error no matter how carefully the base year has been selected.

The simple price and quantity index number are not reliable since the price and quantity of the items are given equal weights. It is obvious that there is irregular and inflationary increase in the prices and quantities and this fact creates a problem of accuracy of the index number of this type.

The merits and demerits of Paasche's and Laspeyre's index make none of them totally acceptable measure of price and volume indices. The Fisher's ideal price index, which is a compromise of Paasche's and Laspeyre's, is not actually perfect because although it satisfies most of the tests, it still has some inadequacies. Statisticians are not relenting in search for a perfect index number.

CLASS ASSESMENT EXERCISE 3

Identity the challenges facing the various forms of index number.

4.0 CONCLUSION

This unit identifies other forms of index numbers such as value index and the consumer price index. The unit equally establishes the relationship between the consumer price index and the inflation rate. The entire unit was rounded up with the problems of index numbers.

5.0 SUMMARY

Value index is the type of index number which takes care of both changes in price and quantity which measures the standard of living over a period of time. The consumer price index is an index that deals with changes in price of goods and services which enables us to obtain the rate of inflation.

None of the measures of index numbers is totally free from one demerit or the other. The merits of one are likely demerits of the other based on the selection of the base value and approaches used in computation. The differences in weight attached and some other irregularities make none of the approaches error free. Hence, experts are working day and night to arrive at an index number approach that is perfect in all ramifications.

6.0 TUTOR MARKED ASSIGMENT

1. Find the value index for the data for seats sold on charter flights.

Year	Price (₦)	No of Seats Sold
1980	85	7,630
1981	88	7,946
1982	93	8,215
1983	98	7,846
1984	99	7,399

Use 1982 as base year.

2. Below is a table of Rental Price Index (Average 1982).

Items	**Indices January 1980 = 100**	**Weight**	
		Civil Servants	**Artisan**
Food	138.2	310	290
Drink	131.8	50	85
Clothing	139.3	70	83
Fuel & light	117.1	140	80
Transport & Vehicle	140.9	60	85
Household durables	101.8	100	90
Snacks	119.4	60	105
Tobacco	120.6	90	67
Services	135.3	95	75
Other goods	130.1	25	40

(a) From the table, compute Retail Price Indices for the Civil Servants and the Artisan.

(b) Estimate the average inflation rate for the economy.

(c) Why do we have difference in the two indices obtained in (i) above.

7.0 REFERENCES / FURTHER READING

Loto, Margaret A, et al (2008): Statistics Made Easy. Concept Publication Limited, Lagos.