NATIONAL OPEN UNIVERSITY OF NIGERIA

COURSE CODE : POL212

COURSE TITLE:

BASIC STATISTICS FOR SOCIAL SCIENCES

COURSE GUIDE

POL212

BASIC STATISTICS FOR SOCIAL SCIENCES

Course Team	Abdul-Rahoof Adebayo Bello (Developer/ Writer) - NOUN
	Adewale Adeyemi James (Developer Writer)
	Dr. (Mrs.) Fadipe-Joseph, O. A. (Editor) - UNILORIN
	Abdul-Rahoof Adebayo Bello (Coordinator) - NOUN
	Prof. A. R. Yesufu (Programme Leader) - NOUN

NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria

Headquarters

14/16 Ahmadu Bello Way

Victoria Island

Lagos

Abuja Office

5, Dar es Salaam Street

Off Aminu Kano Crescent

Wuse II, Abuja

Nigeria

e-mail: @nou.edu.ng

URL: .nou.edu.ng

Published By:

National Open University of Nigeria

First Printed 2010

ISBN: 978-058-488-2

CONTENTS **PAGE**

Introduction

This course is a three-credit unit course for undergraduate students in political science and other social sciences. The materials have been developed with the Nigerian context in view. This course guide gives you an overview of the course. It also provides you with information on the organisation and requirements for the course.

Course Aims

The main aim is to help you have the basic knowledge of statistics as it relates to research in political science. However, the following broad aims will also be achieved:

i) Educating you about the basic concepts and principles of statistics in decision-making process.
ii) Highlighting the uses and limitations of statistics in the society.
iii) Acquainting you with the methodology of data collection and presentation in the political and other social sciences.
iv) To educate you about how to analyse contemporary issues through scientific logical deductions.
v) To enable you understand some basic statistical theories and the importance of measures of central tendencies.

Course Objectives

To achieve the aims set out above, POL212 has overall objectives. In addition, each unit also has specific objectives. The objectives of each unit are stated at the beginning of each unit. It is advisable that you read them before working through the units. Reference may be made to them in the course of studying the units as self-assessments strategy.

Listed below are the wider objectives for the course as a whole. By meeting the objectives, you should be regarded as having met the aims of the course. On successful completion of the course, you should be able to:

a) Discuss the definitions, scope and origin of statistics

b) Explain the approach to data collection

c) Understand the importance and limitation of statistics

d) Explain the tabulation and classification of data

e) Appreciate the graphical and diagrammatical presentations of data

f) Understand the frequency of distribution

g) Understand measures of dispersion and measures of partition define health care financing

h) Explain the probability theory

i) Discuss the Permutation and Combination theory

j) Decipher the binomial distribution

k) Understand the nature and importance of statistical inquiries

l) Appreciate the nature of pure and social science research.

Working through the Course

To complete the course, you are required to read the study units and other related materials. It is also necessary to undertake practical exercises for which you need a pen/pencil, a notebook, graph paper and other materials that will be listed in this guide. The exercises are to aid your understanding of the concepts presented. At the end of each unit, you will be required to submit written assignments for assessment purposes. At the end of the course, a final examination will be written.

Course Materials

The major materials needed for this course are:

(i) Course guide
(ii) Study guide
(iii) Assignment file
(iv) Relevant textbooks including the ones listed under each unit
(v) You may also need to listen to educative programmes and special reports on electronic and print media.
(vi) In addition, read newspapers, news magazine, and academic journals. You also need to interact with computer to explore the Internet facilities.

Study Units

There are 21 units (of four modules) in this course. They are listed below:

Module 1

Unit 1 Definitions, and Scope of Statistics

Unit 2 Approach to Data Collection

Unit 3 Introduction to Set Theory I
Unit 4 Introduction to Set Theory II

Unit 5 Concepts of Logic

Module 2

Unit 1 Diagrammatic Presentation of Data

Unit 2 Frequency Distribution

Module 3

Module 4

Textbooks and References

Certain books have been recommended in the course. You may wish to purchase them for further reading.

Assessment File

An assessment file and a marking scheme will be made available to you. In the assessment file, you will find details of the works that must be submitted to the tutor for marking. There are two aspects of the assessment of this course: the Tutor - Marked Assignment and the written examination. The marks obtained in these two areas will make up the final marks. The assignment must be submitted to the Tutor for formal assessment in accordance with the deadline stated in the presentation schedule and the assignment file.

The work submitted to the Facilitator for assessment will count for 30% of the student's total score.

Tutor-Marked Assignment (TMAs)

You will have to submit a specified number of the TMAs. Every unit in this course has a Tutor - Marked Assignment. You will be assessed on four of them but the best three performances from the TMAs will be used for your 30% grading. When you have completed each assignment, such should be sent together with a Tutor - Marked Assignment Form, to your Tutor. You are advised to make sure that each assignment reaches your Tutor on or before the deadline for submissions. If for any reason, you cannot complete the work on time, contact should be made with the facilitator for a discussion on the possibility of an extension. Extensions will not be granted after the due date unless under exceptional circumstances.

Final Examination and Grading

The final examination will be a test of three hours. All areas of the course will be examined. You should find time to read the unit all over before the examination. The final examination will attract 70% of the total course grade. The examination will consist of questions, which reflect the kinds of self-assessment exercises and Tutor- Marked Assignment previously encountered. Moreover, all aspects of the course will be assessed. You should use the time between completing the last unit, and taking the examination to revise the entire course.

Course Marking Scheme

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

Assessment	**Marks**
Assignments (Best three assignments out of four marked)	30%
Final Examination	70%
Total	**100%**

Presentation Schedule

The dates for submission of all assignment will be communicated to you. You will also be told the date of completing the study units and dates for examinations.

Course Overview

Unit	Title of Work	Weeks Activity	Assessment (End of Unit)
	Course Guide		
	Module 1		
1	Definitions and Origin of Statistics	Week 1	Assignment 1
2	Approach to data collection	Week 1	Assignment 2
3	Introduction to Set Theory 1	Week 2	Assignment 3
4	Introduction to Set Theory 11	Week 2	Assignment 4
5	Concepts of Logic	Week 3	Assignment 1
	Module 2		
1	Diagrammatic Presentation of Data	Week 4	Assignment 3
2	Frequency Distribution	Week 5	Assignment 4
3	Graphical Data Presentation	Week 5	Assignment 5
4	Measures of Central Tendency	Week 6	Assignment 1
5	Measures of Dispersion I	Week 6	Assignment 2
6	Measures of Dispersion II	Week 7	Assignment 3
	Module 3		
1	Probability Theory I		
2	Probability Theory II	Week 9	Assignment 1
3	Permutation Theorem	Week 9	Assignment 2
4	Combination Theorem	Week 10	Assignment 1
5	Binomial Distribution	Week 10	Assignment 2
	Module 4		
1	Nature and Importance of Statistical Inquiries	Week 11	Assignment 1

2	Basic Research Methodology I	Week 12	Assignment 1
3	Basic Research Methodology II	Week 12	Assignment 2
4	Nature of Sciences	Week 13	Assignment 1
5	Some Basic Concepts in Social Statistics	Week 13	Assignment 2
	Total	**Week 13**	

How to Get the Most from This Course

In distance learning, the study units replace the university lectures. This is one of the great advantages of distance learning; you can read and work through specially designed study materials at your own pace, and at a time and place, that suits you best. Think of it as reading the lecture notes instead of listening to the lecturer. In the same way a lecturer might give you some readings to do, the study units tell distance learner where to read, and which are your text materials or set books. Distance

learners are provided exercises to do at appropriate points, just as a lecturer might give you an in-class exercise. 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. Next to this is a set of learning objectives. These objectives will let you know what you should be able to do by the time the units have been completed. These learning objectives are meant to guide you in your studies. The moment a unit is finished, you must go back and check whether you have achieved the objectives. If this is made a habit, then you will significantly 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 a reading section. The following is a practical strategy for working through the course. If you run into any trouble, you should telephone the tutor for assistance. Remember that the tutor's job is to provide such help. You should not hesitate to call and ask for such assistance when needed.

1. Read this Course Guide thoroughly, it is your first assignment.

2. Organise a Study Schedule. Design a 'Course Overview' to guide you through the course. Note the time you are expected to spend

on each unit and how the assignments relate to the units. Whatever method you choose to use, you should decide on and write in your own dates and schedule of work for each unit.

3. Once you have created your own study schedule, do everything to stay faithful to it. The major reason why students fail is that they get behind with their course work. If you get into difficulties with your schedule, let your tutor know before it is too late to help.

4. Turn to Unit I, and read the introduction and the objectives for the unit.

5. Assemble the study materials. You will need your set books and the unit you are studying at any point in time. As you work through the unit, you will know what sources to consult for further information.

6. Keep in touch with your study centre. Up-to-date course information will be continuously available there.

7. Well before the relevant due dates, (about four weeks before due dates) keep in mind that you will learn a lot by doing the assignment carefully. They have been designed to help you meet the objectives of the course and, therefore, will help you pass the examination. Submit all assignments not later than the due date.

8. Review the objectives for each study unit to confirm that you have achieved them. If you feel unsure about any of the objectives, review the study materials or consult your tutor.

9. When you are confident that you have achieved a unit's objectives, you can start on the next unit. Proceed unit by unit through the course and try to pace your study so that you keep yourself on schedule.

10. 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. When the assignment is returned, pay particular attention to your tutor's comments, both on the tutor-marked assignment form and also the written comments on the ordinary assignments.

11. After completing the last unit, review the course and prepare yourself for the final examination. Check that you have achieved the unit objectives (listed at the beginning of each unit) and the course objectives (listed in the Course Guide).

Facilitators/Tutors and Tutorials

Information relating to the tutorials will be provided at the appropriate time. Your tutor will mark 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 take your tutor-marked assignments to the study centre well before the due date (at least 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 if you need help. Contact your tutor if:

- You do not understand any part of the study units or the assigned readings
- You have difficulty with the exercises
- You have a question or problem with an assignment or with your tutor’s comments on an assignment or with the grading of an assignment.

You should try your best to attend the tutorials. This is the only chance to have face-to-face contact with your tutor and ask questions which are answered instantly. You can raise any problem encountered in the course of your study. To gain the maximum benefit from course tutorials, prepare a question list before attending them. You will learn a lot from participating in discussion actively.

Summary

The course guide gives you an overview of what to expect in the course of this study. The course teaches the attitude of the statistical tools used by political and social scientists to analyse issues upon which public policy decisions are based. The course is useful for executives, civil servants, professionals and politicians.

I wish you success in this academic programme, and hope that you will find this course both interesting and useful.

MAIN COURSE

Course Code	POL212
Course Title	Basic Statistics for Social Sciences

Course Team
Abdul-Rahoof Adebayo Bello (Developer/ Writer) - NOUN
Adewale Adeyemi James (Developer Writer)
Dr. (Mrs.) Fadipe-Joseph, O. A. (Editor) - UNILORIN
Abdul-Rahoof Adebayo Bello (Coordinator) - NOUN
Prof. A. R. Yesufu (Programme Leader) - NOUN

NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria
Headquarters
14/16 Ahmadu Bello Way
Victoria Island
Lagos

Abuja Office
5, Dar es Salaam Street
Off Aminu Kano Crescent
Wuse II, Abuja
Nigeria

e-mail: @nou.edu.ng
URL: .nou.edu.ng

Published By:
National Open University of Nigeria

First Printed 2010

ISBN: 978-058-488-2

CONTENTS **PAGE**

MODULE 1

UNIT 1 DEFINITIONS AND SCOPE OF STATISTICS

CONTENTS

1.0 INTRODUCTION

In the modern world of computers and information technology, the importance of statistics is very well-recognised by all disciplines. Statistics has originated as a science of statehood and found applications slowly and steadily in Agriculture, Economics, Political science, Biology, Medicine, Sociology, and Geography etc. To date, there is no other human occupation where statistics cannot be applied. This phenomenon is better captured by the prelude in an excellent literary work on the subject, which states thus:

> "The processing of statistical information has a history that extends back to the beginning of mankind. In early biblical times nations compiled statistical data to provide descriptive information relative to all sorts of things, such as taxes, wars, agricultural crops, and even athletic events. Today, with the development of probability theory, we are able to use statistical methods that not only describe important features of the data but methods that allow us to proceed beyond the collected data into the area of decisions making through generalisations and predictions." (Walpole Ronald E, 1982).

However, to lend credence to this is a presentation in (Johari, J. C., 2005:26) of the behavioural revolution in political science which is contained in Easton's description of its eight 'intellectual foundations' some of which are:

1. **Regularities**: There are certain discernible uniformities in the political behaviour of human beings that can be expressed in generalisations, as they are capable of explaining and predicting social phenomenon.
2. **Verification:** All knowledge must be based on observation and verification. That is, in order to be valid, knowledge should consist of the propositions that may be subjected to empirical investigations.
3. **Techniques:** Correct techniques should be adopted for acquiring and interpreting data, use of research tools or methods, which generate valid reliable and comparative data.
4. **Quantification:** Data should not only be collected, it should also be measured and quantified so that the conclusions of a researcher may be verified based on quantified evidence.

From the above submissions, it is glaring that modern political scientists like to draw conclusions, take decisions and make predictions on political events in a 'scientific' way by sticking to the side of 'facts' (following the techniques of mathematics, statistics, etc.) using sample survey, random sampling, multi-variate analysis, game theory, content analysis, formulation/testing of hypothesis, etc.

2.0 OBJECTIVES

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

- define the term "Statistics"
- give clear and correct examples of statistical data
- understand the role of statistics in political science
- discuss the scope and usefulness of statistics.

3.0 MAIN CONTENT

3.1 Definition of Statistics

The word 'Statistics' has different meanings to different people. To some, it is a collection of tables, charts, data or numbers while to others, it is considered as an aspect of advanced mathematics. However, statistics has become an essential tool in the study and analysis of political science and other social sciences.

Encarta Microsoft Dictionary, 2008, defines the subject as 'a branch of mathematics that deals with analysis and interpretation of numerical data in terms of samples and populations.' According to Wester, (cited in Gupta, 1983), Statistics is a "classified facts respecting condition of the people in a state…..especially those facts which can be stated in numbers or in tables or in any other tabular or classified arrangement." In the same work, Robert W. Bugess summarises what could be regarded as statistics thus:

> The fundamental gospel of statistics is to push back the domain of ignorance, prejudice, rule of thumb, arbitrariness of premature decisions, traditions and dogmatism, and to increase the domain, in which decisions are made and principles are formulated on the basis of analysed quantitative facts. (See Gupta, C. B., 1983:5)

As numerical data and tools for social analysis, 'Statistics' could be r described as measurements, enumerations or estimates of natural phenomenon that are usually systematically arranged, analysed and presented to exhibit important inter-relationships among them.

In everyday affairs, decisions are made based on the available data or information at hand. For example, a politician withdraws from an election as he sampled the opinion of the electorates in his/her constituency and discovered that his chances (probability) of winning the election are slim. Hence, the term "statistics" can be useful as a tool in making effective decisions through a 'scientific' process.

3.2 Some Basic Concepts

i) Data

This could be defined as pieces of information that represent the qualitative or quantitative attributes of a variable or set of variables. Data are typically the results of measurements and can be the basis of graphs, images or observations of a set of variables. Data are often viewed as the lowest level of abstraction from which information and knowledge are derived for statistical analysis.

ii) Variable

This is any quality that can have a number of values, which may be either discrete or continuous. A variable is a property that can take on different values. Individual in a class may differ in sex, age, intelligence, height etc. These properties are variables. Variables could vary in quality or in quantity. *Constants* unlike variables do not assume different values.

iii) Quantitative Variables

This type of variables assumes values that vary in terms of magnitude. Very easy to measure and compare with others e.g. weight, height, age, distance, marks obtained in a test etc.

iv) Qualitative Variables

This type of variable differs in kind. They are only categorised, e.g. gender, nationality, social economic status, academic qualifications, marital status.

v) Independent Variable

These variables can be manipulated or treated. The effect is reflected on the dependent variable. The value of the dependent variable thus depends on that of the independent variable. Note that in graphing, the dependent variable is placed on the vertical (y-axis) while the independent variable is placed on the horizontal (x-axis).

vi) Discrete Variable

This is a variable that can be counted, or for which there is a fixed set of values. For example, the number of votes in an election is a discrete variable.

vii) Continuous Variable

This concept is characterised by being related to some numerical scale of measurement, any interval of which may, if desired, be subdivided into an infinite number of values, e.g. length, height, weight, temperature, volume and time.

viii) Distribution

This is the arrangement of a set of numbers classified according to some properties or attributes such as age, height, weight, etc.

ix) Population

This consists of the totality of the observations of a particular group. For instance, if there are 800 farmers in a community that are engaged in farming, we say the population size is 800. This measurement is of interest representing the aggregate of units to be covered, which could be finite or infinite. When the population can easily be counted then it is said to be finite e.g. the number of contestants for a political post but if the population under consideration is large e.g. the grain of sand, then we say it is infinite.

x) Sample

This is a subset of a population. It is a sub-group or sub-aggregate drawn from a population; i.e. the portion appropriately selected out of the population by the same statistical method for observation.

xi) Parameter

Any numerical value describing a characteristic of a population is known as *parameter*. It is a situation when *mean* (or Average), *standard deviation* or *variance* of a population are computed for statistical analysis.

xii) Statistic

This refers to a descriptive measure of a sample, i.e. a numerical value or function computed to describe a sample or population.

3.3 Types of Statistics

Statistics has three distinct parts namely:

- Descriptive statistics
- Inferential statistics and
- Experimental statistics.

a) Descriptive Statistics

The event or outcome of events is described without drawing conclusions. It is concerned only with the collection, organisation, summarising, analysis and presentation of an array of numerical qualitative or quantitative data. Descriptive statistics include the mean, median, mode, standard deviation, range, percentile, kurtosis, correlation coefficient, proportions etc.

b) Experimental Statistics

Relates to the design of experiments to establishing causes and effects of such designs as experimental, Quasi-experiments etc.

c) Inferential Statistics

This is built on the descriptive statistics by going a step further to make interpretation with a view to population upon which a decision would be based. Valid and reliable decisions, generalisations, predictions and conclusions could be drawn using this statistics tools such as stochastic process, queuing theory, game theory, quality control, chi-square, t-test, f-test etc.

3.4 Functions of Statistics

The roles or functions of statistics are many in the society. The following are few important ones.

3.4.1 Compression

Generally speaking by the word ‘to compress’, we mean to reduce or to condense. This method is applied to facilitate the understanding of a huge mass of data by providing only few observations. If in a particular class of students at the National Open University of Nigeria (NOUN), only marks in an examination are given, no purpose will be served but it would serve a better purpose if we are given the average mark in that particular examination. Similarly, the range of marks is also another measure of the data. Thus, Statistical measures help to reduce the complexity of the data and consequently to understand any huge mass of data.

3.4.2 Evaluation

The two main methods used in condensing data are classification and tabulation method. These help researchers to compare and contrast data collected from different sources. Grand totals, measures of central tendency, measures of dispersion, graphs and diagrams, coefficient of correlation etc provide ample scope for comparison. This is another important function statistics performs. For example, if the rice production (in Tonnes) by the commercial white farmers in Kwara State of Nigeria is known, then we can compare it with the production of the same commodity in Bida, Niger State of the same region or the production of two different regions within Nigeria. A comparative study can be made as statistics is an aggregate of facts, comparison is always possible and in fact, comparison helps us to understand the data in a better way.

3.4.3 Forecasting

By the word forecasting, we mean to predict, estimate or to project into the future. Given the data of the last ten years connected to rainfall of a particular state in Nigeria, it is possible to predict or forecast the rainfall for the near future. In politics, forecasts are possible on voting patterns, election results, etc just as in business where forecasting also plays a dominant role in connection with production, sales, profits etc. The analysis of time series and regression analysis, which are provided by statistics, play a significant role in such exercise.

3.4.4 Estimation

One of the main objectives of statistics is drawing inference about a population from the analysis for the sample drawn from that population. In estimation theory, we estimate the unknown value of the population parameter based on the sample observed. Assuming we are given a sample of heights of hundred students in the School of Arts and Social Sciences of NOUN, based upon the heights of these 100 students, it is possible to estimate the average height of all students in that school.

3.4.5 Tests of Hypothesis

A statistical hypothesis is a statement or postulation or a theory about the relationship between a dependent and independent variables. In the formulation and testing of hypothesis, statistical methods are extremely useful. For instance, we may be interested in knowing whether high rate of unemployment affects re-election of an incumbent President in Nigeria or whether crop yields increases because of the application of new fertilizer or whether the involvement of Emirs in the immunisation campaign is effective in reducing/eliminating polio disease in the Northern part of Nigeria, are some examples of statements of hypothesis and these are tested by proper statistical tools.

3.5 Scope/Uses of Statistics

Statistics is not a mere device for collecting numerical data, but also a means of developing sound techniques for their handling, analysing and drawing valid inferences from them. It is applied in every sphere of human endeavour – social as well as physical sciences – like Biology, Economics, Education, Planning, Politics, Information Technology, etc. It is almost impossible to find a single department of human activity where statistics is not applicable. We now discuss briefly the applications of statistics in other disciplines.

3.5.1 Statistics and Industry

In industries, control charts are widely used to maintain a certain quality level. In production engineering, to find whether the product is conforming to specifications or not, statistical tools, namely inspection plans, control charts, etc., are of extreme importance. In inspection plans, we have to resort to some kind of sampling – a very important aspect of Statistics.

3.5.2 Statistics and Commerce

Statistical data are lifeblood of successful trade and commerce. No business can afford to ignore the inventory or sales records by either over or under stocking of goods. In the beginning the businessperson has to study and estimate, the interplay of market forces (demand, supply and price) for his/her goods and then takes steps to adjust with his output or purchases. Thus, statistics is indispensable in business and commerce. The trend of business adjusts to a number of economic factors. In this connection, market survey plays an important role to exhibit the present conditions and to forecast the likely changes in future.

3.5.3 Statistics and Political Economy

Statistical methods are useful in measuring numerical changes in complex groups and interpreting collective phenomena. Nowadays the uses of statistics are abundantly made in addressing many economic and political problems and it also plays important roles in economic/political theory and practice. Alfred Marshall opines, “Statistics is the straw only which every other economist has to make the bricks.” Statistical tools are immensely useful in solving many political-economic problems such as wages, prices, production, distribution of income, wealth, population census, voting pattern, constituency delimitations and so on.

3.5.4 Statistics and Education

In education sector, the usefulness of statistics cannot be under-estimated since research has become a common feature in all branches of educational activities. Statistics is necessary for the formulation of policies on courses, budget estimation, consideration of facilities available and job creation for the graduates. Many scholars are engaged in research work to test the past knowledge and evolve new knowledge. These are possible only through statistics.

3.5.5 Statistics and Planning

Statistics is indispensable in planning in the modern world. Almost all the ministries, departments and agencies of government are seeking the help of planning for efficient operations, formulation and implementation of policies. In order to achieve this goals, the statistical data relating to all sectors of the state economy and the society at large such as production, consumption, demand, supply, prices, investments, income expenditure etc are collected through statistical techniques for processing, analysing and interpretation. In Nigeria, though not accurately available, the important roles played by statistics in planning both at the central, state and local government levels, cannot be over-emphasised.

3.6 Limitations of Statistics

Since there are no roses without thorns, Statistics with all its seeming bed of roses in every sphere of human activity, have its own limitations and drawbacks some of which are itemised below.

3.6.1 Statistics and the Study of Qualitative Phenomena

Since statistics is a science and deals with a set of numerical data, it is applicable to the study of only these subjects of enquiry, which can be expressed in terms of quantitative measurements. In fact, qualitative phenomenon like honesty, poverty, beauty, intelligence etc, cannot be expressed in terms of number and no statistical analysis can be directly applied on these qualitative phenomena. However, statistical techniques may be applied indirectly by first reducing the qualitative expressions to accurate quantitative terms. For example, the intelligence of a group of students can be studied based on their marks in a particular examination.

3.6.2 Statistics and Individuality

Statistics does not attach any specific importance to the individual items rather; it deals with aggregates of objects. Individual items, when they are taken individually do not constitute any statistical data and do not serve any purpose for any statistical enquiry.

3.6.3 Lack of Exactitude

It is well known that mathematical and physical sciences are exact but statistical laws are not as exact but only approximations. Statistical conclusions may not have universal validity.

3.6.4 Misuse of Records

Statistics must be used only by experts otherwise statistical methods are the most dangerous tools in the hands of the inexperienced people. The use of statistical tools by the untrained persons might lead to wrong conclusions. It may be easily misused by quoting wrong figures of data to dress lies in the gown 'fact' in order to achieve a selfish interest.

SELF ASSESSMENT EXERCISE

What does the concept 'Statistics' connote to you?

4.0 CONCLUSION

Statistics is to social sciences what oxygen is to life and that is the message this unit has tried to convey to you. Statistical methods and approach are necessary tools in the hands of political scientists; analysts and political actors will like to make a mark in the chosen profession. In addition, it is believed that statistics is relevant to all fields of human endeavours, even to students in institutions of higher learning.

5.0 SUMMARY

The subject of statistics embraces collection, organisation, analysis, presentation and interpretation of data for predetermined purposes. For certain reasons, it may not be practicable to obtain data from a large population. Data may therefore be obtained from a sample of a population, analysed using appropriate statistical technique and interpretations are drawn from the results. It is very useful to all spheres of human endeavours but it also has some drawbacks and limitations.

6.0 TUTOR-MARKED ASSIGNMENT

1. What is statistics?
2. What is the importance of statistics to the society?
3. Enumerate the uses and problems of Statistics.

7.0 REFERENCES/FURTHER READING

Allan G. Bluman (2004). *Elementary Statistics*, *A Step by Step Approach*. United States: McGraw Hill Companies.

Boyinbode I.R. (1984). *Fundamental Statistical Methods in Education and Research*. Ile-Ife: DCSS Books.

Gupta, C. B., (1983). *An Introduction to Statistical Method.* New Delhi: VIKAS Publishing House, PVT Ltd.

________. (2004). *An Introduction to Statistical Methods*. New Delhi: Vikas Publishing House PVT Ltd.

Johari, J. C. (2005). *Introduction to Modern Political Principles.*

Walker and Mclean (1973). *Ordinary Statistic*. Edward Arnold Publishers Ltd.

Walpole, Ronald E. (1982). *Introduction to Statistics* (3rd ed.). New York: Macmillan Publishing Co., Inc.

UNIT 2 APPROACH TO DATA COLLECTION

CONTENTS

1.0 INTRODUCTION

By definition, inquiry could mean an investigation to determine the facts of a case. The word could also mean a search for knowledge, but by statistical inquiry, it implies a search conducted according to the statistical technique, which, however, cannot be applied to all kinds of phenomena. Its application is restricted to only those subjects that can be measured quantitatively (Gupta, C. B., 1983:20). Therefore, statistical research into a given problem could be classified as follows:

i) data collection;
ii) organisation of data;
iii) data analysis and
iv) interpretation of facts.

2.0 OBJECTIVES

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

- understand the objects of statistical inquiry
- understand how to plan data collection
- understand the techniques of data collection.

3.0 MAIN CONTENT

3.1 Preliminary Step

The first task a researcher must set for himself /herself is to prepare a statement of purpose of the statistical inquiry s/he is embarking upon. Failure to do this may lead to misunderstanding or confusion, which may result in waste of time, energy and resources.

3.2 Statement of Purpose

In line with the above, a pertinent question, which a researcher should hypothetically ask himself/herself, is: why am I going into this exercise? In an attempt to answer this question, the researcher would be stating the objectives of the statistical inquiry, which could be some or all of the following:

a) To make a new discovery or a breakthrough in a given subject. For instance, you may wish to find a lasting solution to Nigeria's electoral system that has been generally regarded as flawed by coming up with a new system entirely.
b) A social researcher may also wish to know the existing state of affairs such as looking at the Electoral Acts, past and present, with a view to finding out the problem areas.
c) The purpose of a research may also be either to supplement, disprove or test some existing theories or hypothesis in order to determine their suitability or otherwise to the society.
d) It may also be to solve a perennial social problem in the society.

3.3 Plans of Data Collection

The next line of action after the above is for the researcher to draw a plan on how to conduct the investigation/inquiry in which s/he must take cognisance of the following points:

a) The Scope of Inquiry

The coverage area of the inquiry should be determined with reference to:

i) The Space

The general practice as regards the space/scope is to use political/administrative divisions such as a Country, a State, a Local Government Area, a Senatorial District; a Constituency or an electoral Ward. It can also be Economic division such as Agricultural sector; Manufacturing; Mining; Banking or Communication. It can also be social division such as Child Trafficking; Drug Abuse; Prostitution; Armed Robbery, etc.

ii) The Time

The Researcher must know that the findings or outcome of his/her inquiry is time bound and it is a function of how early s/he could collect data. This has to be done within a reasonable time, otherwise, conditions might change and the data collected might not only become useless to the inquiry but would also render the outcome obsolete.

b) Determination of Statistical Unit

This is very important, not only for data collection but also for data analysis, interpretation and presentation. For the gathering of raw materials, clear definition of unit is of primary importance; for the interpretation of the results and the presentation of facts without units is valueless. In order to get a correct diagnosis and solution to a particular problem, the researcher must collect the right data with maximum accuracy and in the appropriate units. For instance, to prevent oil glut that may lead to price instability in the oil market, the oil cartel Organisation of Petroleum Exporting Countries (OPEC) may either reduce members' production quota or reduce members' refinery plants or legislate a maximum price. In this event, the best option is to reduce members' quota otherwise the result would be wrong.

c) Technique of Data Collection

Having determined the scope of inquiry and units of measurement, the next step in drawing up the plan of investigation is to determine the best method of data collection, which may be some of the following measures:

a) **Literature review:** This method involves going through published works by scholars on the subject matter; checking the records of public or organised private organisations that publish data annually, bi-annually or quarterly. This is known as *Secondary source* of data collection.

b) **Survey:** The Researcher may conduct special survey or inquiry on the subject-matter which may include the use of interview or administering questionnaires. This is known as *Primary source*.

Above are the two main sources of data collection and the use of a particular source depends on the following considerations:

i) the purpose of the inquiry;

ii) the time required;
iii) availability of fund;
iv) accuracy required and
v) the nature or status of the researcher.

3.4 What is Primary Data?

This could be defined as the statistical data or materials generated directly by the researcher for the purpose of the research project. As discussed in (b) above, any data collected by an investigator through survey, personal observation; interview; questionnaire; information from correspondents and such other methods are under this category.

3.5 What is Secondary Data?

On the other hand, data or statistical materials collected by investigator from the following printed materials published by another person or organisations are called Secondary data: Newspapers; News magazines; Trade or Academic Journals; Reports; Periodicals, etc. For the purpose of clarity, data collected during census exercise is primary to the National Population Commission but it becomes secondary to every consumer of such data for further research.

SELF ASSESSMENT EXERCISE

What are the plans for data collection?

4.0 CONCLUSION

Whichever method employed by an investigator, it must be borne in mind that objectivity, accuracy of data and thoroughness are the watchwords for a successful and useful research.

5.0 SUMMARY

Today, politics and government activities depend largely, on statistical data and information forecasts of future trends and in order to show the importance of statistical studies, many Nigerian Universities include courses like statistical methods, quantitative politics or political data analysis in their social sciences programmes. Before we proceed in our discussion on the subject of statistics, we consider it fit to discuss some elementary Mathematics such as Sets and Logic, which will aid our understanding of the mathematical aspects of Statistics.

6.0 TUTOR-MARKED ASSIGNMENT

1. Explain the difference between primary and secondary data.
2. Describe the preliminary consideration for data collection.
3. What is the relevance of statistics to political science?

7.0 REFERENCES/FURTHER READING

Gupta, C. B. (1983). *An Introduction to Statistical Methods*. New Delhi: Vikas Publishing House PVT Ltd.

UNIT 3 INTRODUCTION TO SET THEORY 1

CONTENTS

1.0 INTRODUCTION

One of the foundations of Mathematics is the concept of *SETS*. Sets do not serve as the basis of having a good understanding of the subject Mathematics at the lower cadre only, but also, it can be studied as a discipline at higher cognitive level of learning. Credit to the simplicity and the wide real life applications the concept and theory of sets bring to the field of Mathematics as regards its environment. The great German Mathematician George Cantor (1845-1918) defines Sets in a unique and simplified manner, which will be one of our foremost subtopics under this unit.

2.0 OBJECTIVES

At the end of this unit, students will be able to:

- define a set in his/her own understanding
- write in set notation
- solve problems using set notations.

3.0 MAIN CONTENT

3.1 Basic Definitions

3.1.1 Definition of Sets

The great German Mathematician George Cantor (1845-1918) defines a set as any collection of definite distinguishable, objects of our intuition or of our intellect to be conceived as a whole. This highly technical but wholesome definition of *Sets* has being simplified over the years as a collection of well-defined objects or better still, collection of objects.

Here are some examples of sets:

1. A collection of students in a school.
2. Letters of the alphabet.
3. The numbers 2, 3, 5, 7, and 11.
4. A collection of all positive numbers.
5. The content of a person's purse.
6. A collection of books in the library
7. The human race
8. The state capital of a country is a set.
9. A collection of historical artefacts is a set, etc

3.1.2 Definition of Real Numbers

We denote real numbers by $\mathbb{R}$, the set of numbers 1,0,2,-2,1/2, -3/8, $\pi, \sqrt{5}$, etc., these are what we are familiar with from elementary school and they are called Real Numbers. The subsets of $\mathbb{R}$ are listed below with examples

$\mathbb{Z}$= the integers…, -3, -2, -1, 0,1,2,3,…
$\mathbb{N}$ = *the natural numbers (also called the positive integers)* … , 1,2,3 …
$\mathbb{Q}$= the rational numbers of the form $\frac{a}{b}$ where a and b are integers and b$\neq$ 0

3.1.3 Definition of Members of a Set

We know that any clearly defined collection of things, objects or numbers makes up a set. Therefore, each member of a set is called an element of the set.

3.2 Set Notation

A set is usually denoted by capital letters (X, Y, Z…) while the elements or members of the set are written with small letters (x, y, z…). These objects are called members or elements of a set.

The correct representation of a set is to write the elements, separated by commas and enclosed between braces or curly brackets.

e.g., Set *A = {a, b c, d, e, f, g}.*

3.3 Membership of a Set

The statement "c" is an element or member of set *A* or "c" belongs to "*A*" is written as $c \in A$. The contrary statement that "c" does not belong to *A* is written as: $c \notin A$.

There are two ways of specifying a set. One way is by listing the elements in the set such as:

$A = \{a, b, c, d\}$. If Y is the set of prime numbers and k= 7 then k $\in Y$ and if c = 8 then $c \notin Y$

A second way of describing a set is by stating the rule or property, which characterizes the set.

For example, $B = \{x / 3 < x < 6\}$ or $B = \{x : 3 < x < 6\}$. Notice, the stroke / or colon: can be used interchangeably, which means 'such that' the representation $B = \{x / 3 < x < 6\}$ is read as follows: *B* is a set consisting of elements *x*, such that 3 is less than *x* and *x* is less than 6.

3.3.1 Tabular Form and Builder Form of a Set

If a set is specified by listing its elements, we call it the *tabular form* of a set;

If it is specified by stating its property, such as $C = \{x / x \text{ is odd}\}$, then it is called *builder form of a set.*

In general a set can be specified in the following ways

- By listing all the members of the set
- By describing the elements of the set
- By using braces to enclose every element that can be associated with that set.

3.4 Finite and Infinite Sets

3.4.1 A Finite Set

A finite set is one whose members are countable.

Examples of finite sets are given below:

i. The set of students in a school
ii. The contents of a lady's hand-bag;
iii. Whole numbers between 10 and 100,000.
iv. Members of a football team.
v. Members of the state house of assembly
vi. Governors of Federal Republic of Nigeria.

3.4.2 An Infinite Set

An infinite set is one whose elements are uncountable, as they are infinitely numerous.

Examples of the infinite sets are listed below:

i. Rational numbers
ii. Real numbers.
iii. The set B = $\{all\ positive\ integers\}$
iv. Complex numbers.
v. Positive even numbers.

The main distinction between a finite set and the infinite set is that a finite set has a definite beginning and a definite end, while the infinite set may have a beginning and no end or may not have both beginning and end.

For example, we specify the set of positive even numbers, as follows:

$P = \{2,4,6,8,10...\}$ or
$P = \{x : x > 2, x \text{ is even}\}$

3.5 Subset and Equality of Sets

Let $A = \{a,b,c,d,e,f\}$ and $B = \{c,d,e\}$, then we say B is contained in A, and we use symbol '$\subset$' to denote the statement 'is contained in', or 'is a subset of'. Thus $B \subset A$, is read as 'B is contained in A'; if there is an *x*, such that $x \in B$ implies $x \in A$. The statement B is contained in *A* can be put in reverse order as 'A contains B' and we will write $A \supset B$. However, this form is not very popular. If B is not a subset of set $C = \{3,4,a\}$, then we write $B \not\subset C$. It should be noted that unless every member of B is also a member of A, then could we say B is a subset of A.

Example

If A={2,4,6,16,8,10,12,14}, B={4, ,2,12}, C={6,10} D={2,4,6,8,10,12}
Note that the order of the element in a set does not matter. From the above example we can draw the following conclusions

1. $B \subset A$
2. $C \subset A$
3. $D \subset A$
4. $A \supset B$

Two sets A and B are said to be equal if and only if $A \subset B$ and $B \subset A$. Suppose $A = \{1, 2, 3\}$ and $B = \{3, 1, 2\}$ then, A=*B*. Note that the rearrangement of the elements of a set does not alter the set.

3.6 Types of Sets

3.6.1 Empty or Null Set

A set that contains no element is called null or empty set. Null means void, therefore, a null set is an empty set. The null set is denoted by the symbol ø. Alternatively, we can use a pair of open and closed braces { } to denote an empty set instead of ø. Note however that {0} cannot be classified as a null set, because it has an element, zero.

Example 1

Given that set C=$\{Days\ of\ the\ week\ that\ begins\ with\ letter\ z\}$
Hence C= $\{\quad\}, or\ \Phi$

Example 2

Given that set B=$\{Square\ with\ 10\ sides\}$
Hence B= $\{\quad\}, or\ \Phi$

3.6.2 The Cardinality of a Set

The number of elements in a set is called its cardinality. This is usually denoted by n(A) where A is a finite set.

Example

If A = $\{a, b, c, d, e, f, g, h\}$ then n(A) = 8

3.6.3 Singleton

Any set, which has only one element, is called a singleton.

Examples

A= {a} is a singleton.
n(A) = {2} implies that n(A) = 1

3.6.4 The Universal Set

Universal set is the set of all elements under consideration. Every set is a subset of a larger or equivalent fixed set. This larger set is called the universal set. In other words, the total collection of elements under discussion is called the universal set.

The symbol μ or E is often used to denote a universal set. For example, if we toss a die, once, we expect to have either a "1" or a "2" or a "3" or a "4" or a "5", or a "6" as a result. If there are no other expected results different from these numbers, then we say, for this particular experiment, the universal set is {1, 2, 3, 4, 5, 6,}. Thus a universal set is the total population under discussion.

3.6.5 Proper Subsets

If A is a subset of B and if there is at least one member of B which is not a member of A, then A is a proper subset of B and we write $A \subseteq B$.

Consider the set A={1, 2, 3}. The following sets {1, 2, 3}, {1, 2}, {1, 3,}, {2, 3}, {1}, {2}, {3}, ø are proper subsets of A. Thus {1, 2, 3} $\not\subset$ {1, 2, 3}, but {1, 2, 3} $\subseteq$ {1, 2, 3}.

3.6.6 Power Set

The collection of all the subsets of any set *S* is called the power set of *S*. If a set has *n* members, where *n* is finite, then the total number of subsets of *S* is 2^n. Occasionally we denote the power set *S* by 2^s.

Example

Let P = {*a*,*b*,*c*}

The subsets of P are: {*a*,*b*,*c*}, {*a*,*b*,}, {*a*,*c*}, {*b*,*c*}, {a}, {b}, {c} and ∅.
The power set of A written p(A) = 2^3 subsets; as seen above. Conventionally ∅ is a subset of every set, and it is a proper subset of every set except itself. The symbol (is often used loosely to denote subset and proper sub-set.

Example

Find the power set 2^S of the sets

a. S = {3, 4}
b. P = {a, {1, 2}}

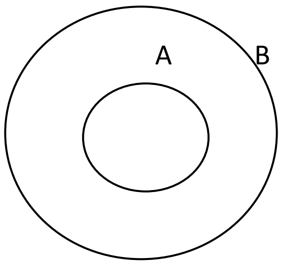
Solution

a. 2^s = {3, 4}, {3}, {4}, ø since the n(S) is 2 hence we have 2^2 number of subsets
b. 2^p = {a, {1, 2}}, {a}, {1,2}, ø

3.7 Venn – Euler Diagrams

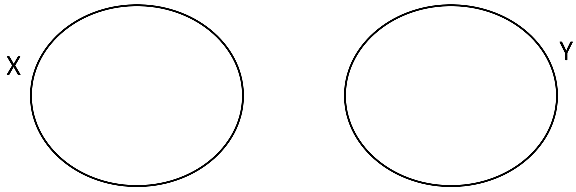
Set theory can be better understood if we make use of the Venn-Euler diagrams. The Venn-Euler diagram is an instructive illustration, which shows relationships between sets.

Suppose $A \subset B$ and $B \neq A$ we can represent this statement in a Venn-Euler diagram as follows:



3.8 Disjoint Sets

Two sets X and Y are said to be disjoint if neither X nor Y has elements in common. The Venn diagram for disjoint sets X and Y is the diagram below.



3.9 Set Operations

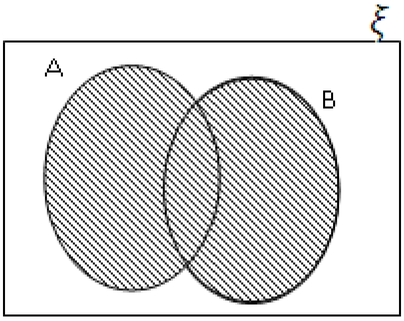
In set, we use symbols ∪ read 'union' and ∩ read 'intersection' as operations. These operations are similar but not the same as the operations in arithmetic. At the end of this chapter, a reader of this topic should be able to identify areas of analogy between the operation in arithmetic and those of a set.

3.10 Union of Sets

The union of sets A and B is the set of all elements which belong to A or to B or to both A and B. This is usually written as $A \cup B$, and reads 'A union B'.
In set notation, we define $A \cup B$ as:

$$A \cup B = \{x : x \in A \; or \; x \in B\}.$$

The shaded portions in the Venn diagram

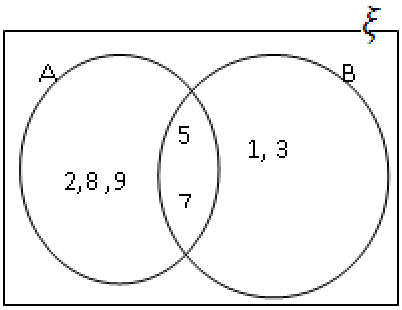


Example

Find the union of Sets A and B if A= $\{2,5,7,8,9\}$ and B= $\{1,3,5,7\}$

Solution

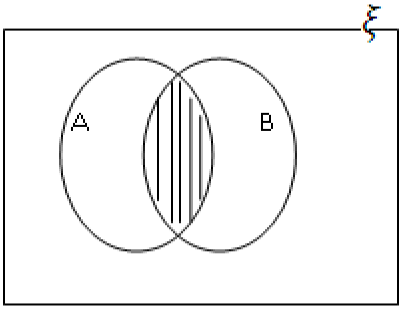
A$\cup$ B= $\{1,2,3,5,7,8,9\}$



3.11 The Intersection of Sets

The intersection of sets A and B is the set of elements which belong to both A and B. Simply, 'A intersection B', written $A \cap B$ consists of elements which are common to both A and B.

The Venn-Euler diagram which represents $A \cap B$ is the shaded portion.



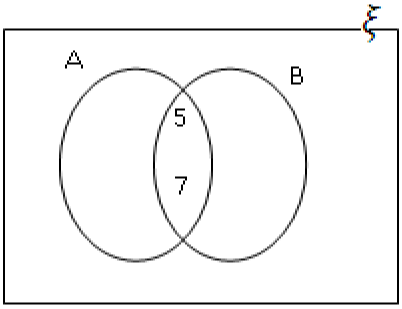
In set notation,

$$A \cap B = \{x : x \in A \text{ and } x \in B\}$$

Example

Find the intersection of the sets A= $\{2,5,7,8,9\}$ and B = $\{1, ,3,7,5\}$
A∩ $B = \{5,7\}$

This can be represented in a Venn diagram



Examples

Let μ= $\{1,2,3,4, - - - 10\}$
A = $\{odd\ \ numbers\ up\ to\ 9\}$
B = $\{Numbers\ less\ than\ 7\}$

Write out the members of the following

a) A∪B
b) A∩B
c) A∪ μ
d) A∩ μ

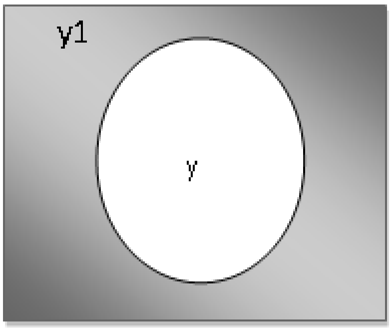
Solutions

Given $\mu = \{1,2,3,4,5,6,7,8,9,10\}$
A=$\{1,3,5,7,9\}$
B=$\{0,1,2,3,4,5,6,\}$

a) $A \cup B = \{0,1,2,3,4,5,6,7,9\}$
b) $A \cap B = \{1,3\}$
c) $A \cup \mu = \{1,2,3,4,5,6,7,8,9,10\}$
d) $A \cap \mu = \{1,3,5,7,9\}$

3.12 The Complement of a Set

Complement of a set y is the set of elements, which do not belong to y, but belong to the universal set(μ). The complement of a set y is usually represented by y' or y^c. The complement of y is represented in the Venn-Euler diagram as shown below



In set language, $A^c = \{y : y \in \mu, or\ y \notin A\}$

Example

Given that A, B and C are subsets of the universal set μ, each of which is defined as follows:

$$\mu = \{x : 2 \le x < 12, x \text{ is an integer}\}$$
$$P = \{x : 3 < x < 6\}$$
$$Q = \{x : (2 < x \le 5) \cup (9 < x < 12)\}$$
$$R = \{x : 4 \le x \le 8\}$$

a. List the members of sets μ, P, Q and R.

b. Find

(i) $(P \cup Q) \cup R$

(ii) $P \cup (Q \cap R)$

(iii) $P \cap (Q \cup R)'$

Solution

a. $\mu = \{2,3,4,5,6,7,8,910,11\}$

$P = \{4,5\}$

$Q = \{3,4,5,10,11\}$

$R = \{4,5,6,7,8\}$

b. (i) $P \cup Q = \{3,4,5,10,11\}$

$\therefore \quad (P \cup Q) \cup R$

$= \{3,4,5,6,7,8,10,11\}$

(ii) $Q \cap R = \{4,5\}$

$\therefore \quad P \cup (Q \cap R) = \{4,5\}$

(iii) $Q \cup R = \{3,4,5,6,7,8,10,11\}$

$(Q \cup R)' = \{2,9\}$

$\therefore \quad P \cap (Q \cup R)' = \Phi$

Example

a. Given the universal set $\mu = \{1,2,3,4,5\}$ and $P = \{1,2,4\}$ $Q = \{2,4,5\}$. Find $P' \cap Q$.

b. If μ is the universal set consisting of all positive integers and *P,Q, R* are subsets such that
P = {*x:x* is a prime number}.
Q = {*x:x* is an even number}
R = {*x:7*< *x* ≤20}.

List the elements of:

(i) $P \cap R$

(ii) $Q \cap R$

(iii) $P' \cap (Q' \cap R)$

Solutions

a. $P' = \{3,5\}$

$\therefore \quad P' \cap Q = \{5\}$

b. The set $P = \{2,3,5,7,11,13,17,...\}$

$Q = \{2,4,6,8,10,...\}$

and $R = \{8,9,10,11,12,13,14,15,16,17,18,19,20\}$

(i) $\therefore \quad P \cap R = \{11,13,17,19\}$

(ii) $Q' = \{1,3,5,7,9,...\}$

$\therefore \quad Q' \cap R = \{9,11,13,17,19\}$

(iii) $P' \cap (Q' \cap R) = \Phi$

SELF ASSESSMENT EXERCISE

1. Given that the universal set

$\mu = \{a,b,c,d,e,f,g,h,i,j,k\}$

$A = \{a,d,e,i,k\}$

$B = \{f,g,i,j,k\}$

using the above sets, show that

$(A \cup B)^1 = \{A^1 \cap B^1\}$

2. If the universal set μ is given by

$\mu = \{a,b,c,d,e,f,g,h\}$

and the sets

$A = \{a,d,c\}$

$B = \{b,c,d,f,h\}$

$C = \{b,c,d,f,h\}$

Find

i. $A \cup (B \cap C)$

ii. $(A \cap B)'$

4.0 CONCLUSION

Set Theory is a very important and fundamental concept in the field of Mathematics and a good understanding of set theory will enhance the study of other Mathematics related fields of study such as Statistics.

Every set is a subset of itself and a subset of the universal set. In everyday language, when numbers are uncountable, they are said to be infinite or unbounded. On the other hand, when a set can be counted and the counting has a definite beginning or end, then we are dealing with a finite set.

A set is a collection of objects. Every set is usually enclosed between curly brackets. Every set is usually written with capital letters, while the members in a set are written with small letters.

5.0 SUMMARY

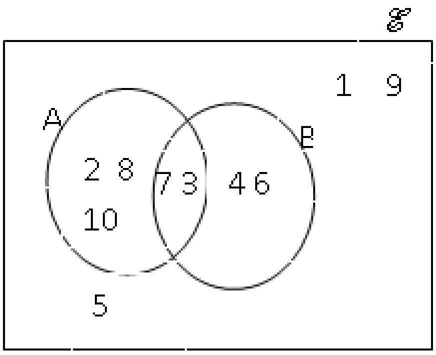
In this unit, we have considered the fundamentals of set theory, we started by considering the definitions of some basic terms in set theory and we looked at the following concepts in set theory.

1. Two sets A and B are said to be equal if $A \subset B \text{ and } B \subset A$.
2. The Venn-Euler diagram is an instructive illustration, which portrays the relationship between sets.
3. The union of two sets A and B is the set of all elements of set A or of set B or both, and we write $A \cup B$.
4. The intersection of sets A and B consists of elements common to A and B and we write $A \cap B$.
5. The complement of a set *x* is the set of all the elements which do not belong to *x*, but belong to the universal set.

6.0 TUTOR-MARKED ASSIGNMENT

1. What is the union of $\{January, February, March\}$ and $\{April, May, June\}$?

   What is their intersection? Represent the two sets in a Venn diagram.
2. Below is a Venn diagram representing a universal set and subsets A and B.



List the elements of the following sets

a) $A\cup B$ b) $A\cap B$ c) $E\cap B$ d) $A\cup E$

3. Which of the following pairs of sets are disjoint?

a) $\{even\ numbers\}\ and\ \{odd\ numbers\}$
b) $\{houses\ in\ Afica\} and\ \{houses\ in\ Ibadan\}$
c) $\{rivers\ in\ Africa\} and\ \ \{Thames, mississippi, Niger\}$
d) $\{letters\ in\ pupil\}\ and\ \{letters\ in\ teacher\}$

7.0 REFERENCES/FURTHER READING

Daniel W.W and Terrel J.C. (1979). *Business Statistics; Basic concepts and Methodology* (2nd ed.). Boston: Houghton Mifflin Co.

Harper W. M. (1990). *Statistics for Management* (4th ed.). Macdonald and Evans Handbook Series.

UNIT 4 INTRODUCTION TO SET THEORY II

CONTENTS

1.0 INTRODUCTION

The concept of set is very important because set is now used as the official mathematical language. A good knowledge of the concept of set is, therefore, necessary if Mathematics and Statistics are to be meaningful to their users. In this unit, we will build on the fundamentals of sets theory that were discussed in the previous unit. However, our focus will be to consider some algebra of set and its properties with practical examples that will aid the understanding of the course in view.

2.0 OBJECTIVES

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

- state the algebraic properties of sets
- define and state the importance of number sets
- solve problems using set notation
- solve problems using Venn diagram.

3.0 MAIN CONTENT

3.1 The Algebra of Sets

The operations of union $\cup$ and intersection $\cap$ are loosely analogous to those of addition and multiplication in algebra. Without any loss of generality, we can safely apply the laws of algebra to sets. The venn diagram provides a method for showing certain results in algebra of sets. We will use an example to illustrate this.

Example

If A, B and C are non-empty subsets of a Universal set μ, show by means of venn diagram that:

a. $(A \cap B) \cap C = A \cap (B \cap C)$

b. $(A \cup B) \cup C = A \cup (B \cup C)$

c. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Solutions

a) $(A \cap B) \cap C = A \cap (B \cap C)$?

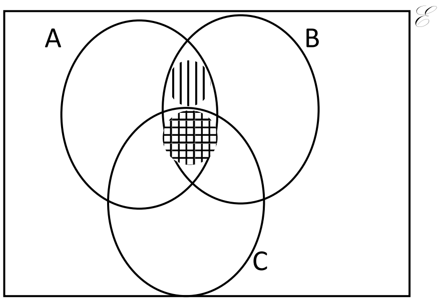


Fig. 1

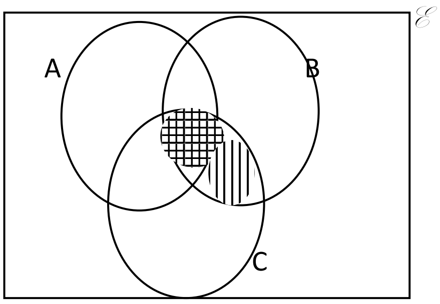


Fig. 2

In fig 1: $(A \cap B)$ is the portion shaded vertically and the portion shaded horizontally and vertically is $A \cap (B \cap C)$

In fig 2: $(B \cap C)$ is the portion shaded horizontally. $A \cap (B \cap C)$ is the portion shaded horizontally and vertically. Since the double shaded portions in the figures are the same therefore $(A \cap B) \cap C = A \cap (B \cap C)$.

b) (AUB)UC= AU(BUC)

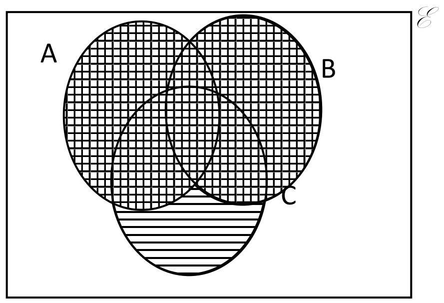


Fig 3

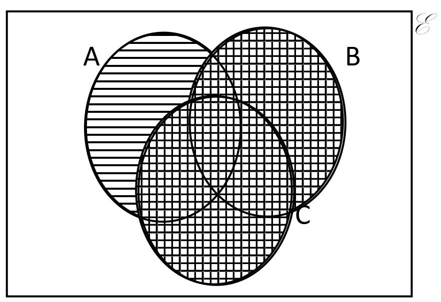


Fig 4

In fig. 3: AUB is the portion shaded vertically. AU (BUC) is the portion shaded either horizontally or vertically or both horizontally and vertically. In fig 4 (BUC) is the portion shaded vertically. AU (BUC) is the portion shaded horizontally or vertically or both. Since the shaded portions in the two figures are the same, therefore
(AUB)UC = AU(BUC)

c) $A\cup(B\cap C) = (A\cup B)\cap(A\cup C)$

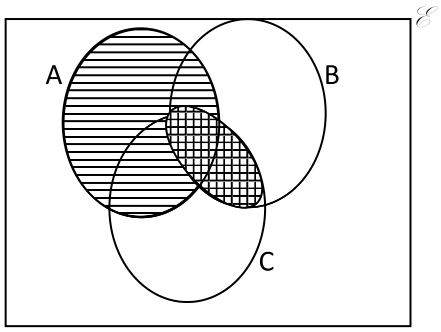


Fig. 5

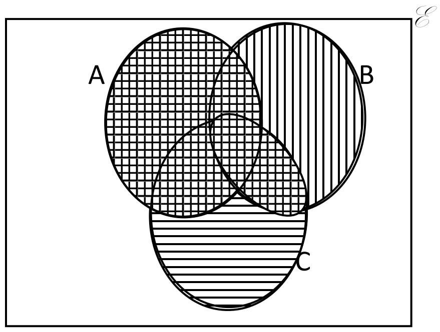


Fig. 6

In fig. 5: $(B\cap C)$ is the vertically shaded portion. While AU $(B\cap C)$is the portion shaded horizontally or vertically or both. In fig. 6: (AUB) is the portion shaded vertically. AUC is the portion shaded horizontally, (AUB) $\cap$(AUC) is the portion shaded both vertically and horizontally. Hence, since the shaded portion in fig. 5 is the same as the doubly shaded portion in fig. 6. It therefore follows that
$A\cup (B\cap C) = (A\cup B)\cap(A\cup C)$

SELF ASSESSMENT EXERCISE 1

Use Venn diagram to prove the following:

a. $A\cap(B\cup C) = (A\cap B)$ U $(A\cap C)$

b. $(A\cap B)' = A'\cup B'$

c. $(A\cup B)'=A'\cap B'$

3.2 Closure Property

If A and B are sets, which are subsets of the universal set μ then the following
g hold:

$A\cup B\subset\mu$ and $A\cap B\subset\mu$. The analogy in algebra; using the operations of + and X are 3 +4 =7 $\in\mathbb{R}$ and 3 X 4 = 12 $\in\mathbb{R}$ is the real number system. If the addition or multiplication of 3 and 4 gives some number that cannot be found in the real number system $\mathbb{R}$, we say the operation of + or X is not closed. Similarly, in set theory, the operations of union and intersection are closed.

3.3 The Commutative Law

$X\cup Y=Y\cup X$ and $X\cap Y=Y\cap X$. Parallel examples in arithmetic are 2 + 3 = 3 + 2 and 2 X 3 = 3 X 2. Thus any two set are commutative with respect to $\cup$ *and* $\cap$.

3.4 The Associative Law

$X\cup(Y\cup Z)=(X\cup Y)\cup Z$ and
$X\cap(Y\cap Z)=(X\cap Y)\cap Z$.

Again, sets obey the associative law of algebra.

3.5 The Identity

In every day arithmetic, 0+1= 1+0=1 and 2 X 1 = 1 X 2 = 2 are two correct solutions. The '0' (zero), in the first case is called the *additive identity*; while '1' in the second case is called the *multiplicative identity*. By a similar analogy, every set has quantities Ø and µ with the property that:

(i) $X\cup\emptyset=\emptyset\cup X=X$

(ii) $X\cap\mu=\mu\cap X=X$

Thus Ø is the identity with respect to union $\cup$ and µ is the identity with respect to intersection $\cap$.

3.6 Inverse

In the set of real numbers R, a + (-a) = (-a) +a = 0 and a x a^{-1} = 1. Thus a number *x* operated on its inverse gives identity.

Similarly, in set theory, every set has an inverse with respect to the operations of $\cup$ and $\cap$ such that:

(i) $X \cup X' = X' \cup X = \mu$ and
$X \cup \mu = \mu \cup X = \mu$

(ii) $X \cap X' = X' \cap X = \emptyset$ and
$X \cap \emptyset = \emptyset \cap X = \emptyset$

SELF ASSESSMENT EXERCISE 2

Show the above identities using Venn diagrams.

3.7 The Distribution Law

$X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$ and
$X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$.

The operation of union is distributive over the operation of intersection and vice versa.

3.8 The Laws of Complementation

(i) $X \cup X' = \mu$

(ii) $(X')' = X$

(iii) $(X \cup Y)' = X' \cap Y'$

(iv) $(X \cap Y)' = X' \cup Y'$

Items (ii) – (iv) are called De-Morgan's laws.

3.9 Practical Examples

Use Venn-Euler diagram to verify that

$$n(A \cup B \cup C) = n(A) + n(B) + n(C)$$
$$- n(A \cap B) - n(A \cap C)$$
$$- n(B \cap C) + n(A \cap B \cap C)$$

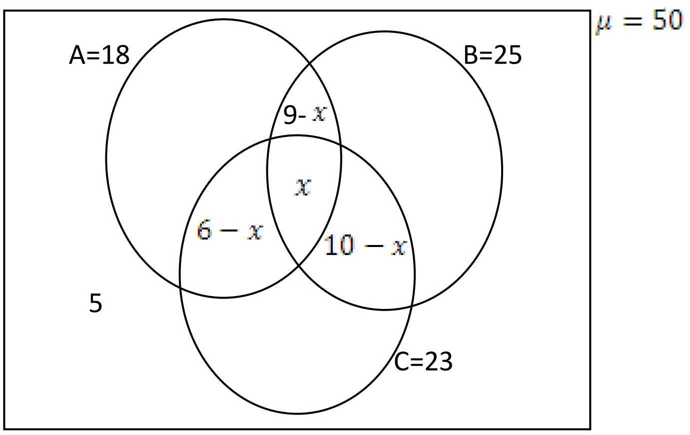
Examples

During the 1st semester, in the Department of Political Science, National Open University of Nigeria, 18 students took POLs. 101, 25 took POLs 102, 23 took POL 103, 9 took POL 101 and POL 102; 10 took POL 102 and POL 103; and 6 took POL 101 and POL 103. If there were 50 students and 5 students did not take any of three courses. How many students took:

i. All three courses?
ii. Only POLs 102?
iii. POLs 103 but not POLs 102?
iv. POLs 101 and POLs 103 but not POLs 102?

Solutions

Let A, B, and C denote the students who took POLs 101, 102, and 103 respectively in the following Venn-Euler diagram. Let x denote those who offer all three subjects and μ, the universal set.



Form the Venn-Euler diagram,

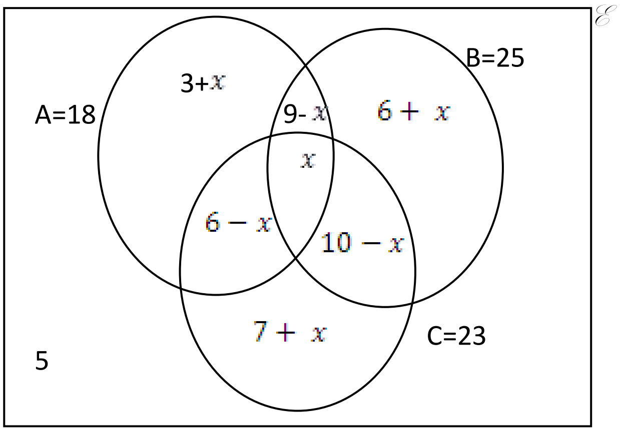
$n(B \cap C \cap A') = 10 - x$
$n(A \cap B \cap C') = 9 - x;$
$n(A \cap C \cap B') = 6 - x$
Hence

$n(A \cap B \cap C') = [18 - (9 - x) - (6 - x) - x] \quad = 3 + x$

Meaning the set of those who read only POLs 101 but not POLs 102 and not POLs 103. Also $n(B \cap A' \cap C') = 25 - (9 - x) - (10 - x) - x = 6 + x$

Similarly $n(C \cap A' \cap B) = 23 - (6 - x) - (10 - x) - x = 7 + x$

We can now draw another Venn diagram and insert these new quantities, as shown below



The total number of students who took 1 or 2 or 3 courses = 50 – 5 = 45.

Thus,
45 = (3 + x) + (9 – x) + x + (6 – x)+(6 + x) + (10- x) + (7 + x)
45 = 41 + x
x = 4

i. Thus, those who took all three courses were 4 in number.
ii. Those who took only POLs 102 were 10
iii. Those who took POLs 103 but not POLs 102 were 13 in number.
iv. Those who took POLs 101and POLs 103 but not POLs 102 were 2 only.

Example

A survey of the number of university students who read three newspapers was conducted. Their preferences for the New Nigerian, the Daily Times and the Pioneer are given below:

48% read the New Nigerian,
26% read the Daily Times,
36% read the Pioneer,
8% read the New Nigerian and Pioneer
5% read the Daily Times and Pioneer,
4% read the New Nigerian and the Daily Times, while
4% read none of the newspapers.

Calculate the percentage of those who read:

i. All three newspapers;
ii. Only the Pioneer,
iii. Only the New Nigerian.

Solution

i. Let
N = Students who read the New Nigerian.
D = Students who read the Daily Times
P = Students who read the Pioneer.

Suppose 100 students were interviewed.

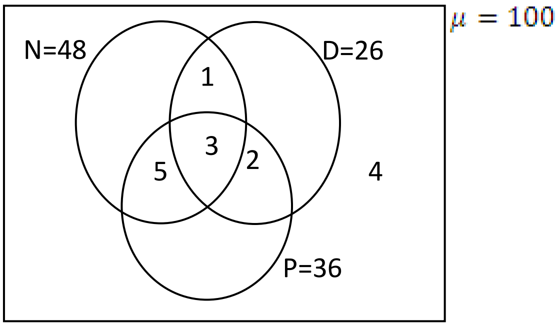
Then

$n(N) = 48, n(D) = 26, n(P) = 36, n(N \cap P) = 8, n(D \cap P) = 5, n(N \cap D) = 4$ Since 4% of the students read neither N, D nor P, then $n(N \cup D \cup P) = 96.$

Hence, by the formula

96 = 48 + 26 + 36 – 8 – 5 – 4 + $n(N \cap D \cap P)$
96 = 93 + $n(N \cap D \cap P)$
$\therefore n(N \cap D \cap P) = 3$

Hence 3% of the students read all three newspapers.



ii. Since $n(N \cap D \cap P)$= 3 and $n(D \cap P) = 5$
$\therefore n(N \cap D \cap P) = 5 - 3 = 2$
Also $n(N \cap D' \cap P) = 8 - 3 = 5$
$\therefore n(P) = 36 - 5 - 3 - 2 = 26\%$

iii. Thus those who read only the New Nigerian = 48 – 5 -3 – 1= 39%

4.0 CONCLUSION

The Venn diagram is an instructive illustration, which portrays the relationship between sets. Sets obey the algebraic laws of closure, the commutative, the associative and the distributive laws. Each set has an identity and an inverse with respect to the operations of union and intersection.

5.0 SUMMARY

In this unit, we have considered the following:

1. Sets obey the algebraic laws of closure, the commutative, the associative and the distributive laws. Each set has an identity and an inverse with respect to the operations of union and intersection.
2. We have looked at practical examples on set theory
3. Each finite set can be counted and the number of elements which constitute it can be written in the manner: n(A) = 20, means set A has 20 elements.
4. Also, if three sets are not disjoint then

$$n(A \cup B) = n(A) + n(B) + n(C)$$
$$- n(A \cap B) - n(A \cap C)$$
$$- n(B \cap C) + n(A \cap B \cap C)$$

6.0 TUTOR-MARKED ASSIGNMENT

1. In a class of 50 students, 25 offer Mathematics, 22 offer Physics, 30 offer Chemistry and all the students take at least one of these subjects. 10 offer Physics and Mathematics, 8 Offer Chemistry and Mathematics, 16 offer Physics and Chemistry.

a) Draw a Venn-Diagram to illustrate the information
b) Find the number of students who offer all the three subjects

2. Show by use of appropriate example if $A \subset B \text{ and } B \subset C \text{ then } A \subset C$.
3. A Platoon of Soldiers (32 in number) went for a night raid in an enemy location. On their return, it was found that 12 were shoot in the head, 16 in the hand and 10 in the leg. 8 had wound on the head and hand, 4 were wounded in hand and leg and 6 in leg and hand. Assume all soldiers were wounded as above, describe, how many were shot in all the three parts of the body?

7.0 REFERENCES/FURTHER READING

Daniel W.W and Terrel J.C. (1979). *Business Statistics; Basic concepts and Methodology* (2nd ed.). Boston: Houghton Mifflin Co.

Harper W.M (1990). *Statistics for Management* (4th ed.). Macdonald and Evans Handbook Series.

Levin R.I. (1990). *Statistics for Management* (4th ed.). New Delhi: Prentice Hall of India Private Ltd.

Stevenson W. J. (1978). *Business Statistics: Concepts and Application*. New York: Harper and Row Publishers.

UNIT 5 CONCEPTS OF LOGIC

CONTENTS

1.0 INTRODUCTION

Logic is the science of reasoning and thinking so that suitable inferences and conclusions after persuasive argument can be made. In other words, logic can simply be defined as the study of arguments. The concept of logic is very important to any Social Science researcher; it helps to draw conclusions when faced with varying data or information from respondents.

2.0 OBJECTIVES

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

- distinguish and define statements and non-statements
- use logic notations to represent statements
- list and explain the two types of statement e.t.c.
- write the negation of a statement
- write the converse of a statement
- write the inverse of a statement.

3.0 MAIN CONTENT

3.1 Concepts of Argument and Statement

An argument is a collection of statements specially arranged. A statement (in logic context) is a sentence that is either TRUE or FALSE but NOT both simultaneously.

The following are statements:

a) Ilorin is in Nigeria
b) $8 > 5$
c) This is the senate building of my university
d) Toyosi is a member of my club
e) It is raining
f) $2 + 1 = 5$

The following are NOT statements:

g) Come to my office
h) How are you today?
i) Look at me
j) Sky is your limit
k) Do you understand?
l) Who are you?

Generally, questions, exclamations, commands, and expressions of feelings, which can be satisfactorily assigned a truth-value, are NOT statements in the logical context.

3.2 Notations in Logic

A true statement is said to have a truth value T while a false statement is said to have a truth value F. By convention, we shall use small letters e.g. p, q, r, s, t, etc to denote statements.

3.3 Types of Statements

There are two basic categories of statements in Mathematics namely: simple statement and compound statement.

3.3.1 Simple Statement

A simple statement is a statement that cannot be further broken down into simpler statements. All the examples (a) to (f) are simple statements.

3.3.2 Compound Statements

A compound statement is a combination of two or more simple statements.

Example

"This is 100 level class" and "Toyosi is a member of the class" is a compound statement also called Composite *Statement*.

There are many ways of connecting simple statements to form compound statements. Some of these ways will be examined.

3.3.2.1 Conjunction (AND) Connective

This connective is usually denoted by the symbols $\wedge$ placed between two simple statements such as statement $p \wedge q$ read as "p AND q".

Example,

Let p = "This is 100 level class"
q = "Toyosi is a member of the class"

Then $p \wedge q$ = "This is 100 level class AND Toyosi is a member of the class".

The truth value table for $p \wedge q$ is given as:

P	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

Note: The statement $p \wedge q$ is only true when the two simple statements are both true.

3.3.2.2 Disjunction (OR) Connective

This connective is usually denoted by the symbol v placed between simple statements such as statement p v q read as "p OR q".

Example

Let: p = " x + 5 > 2"
q = "y + 6 < 14"

Then, p v q = " x +5 > 2" OR "y + 6 < 14"

The truth table for p v q is given as:

P	q	pvq
T	T	T
T	F	T
F	T	T
F	F	F

Note: The statement p v q is only false when the statements p and q are both false.

3.4 Bi Conditional Statement (If and Only If)

This connective is usually denoted by the symbol $\Leftrightarrow$ placed between simple statements such as statement p $\Leftrightarrow$ q read as: "p if and only if q" meaning that p is a sufficient condition for q".

Example

Let p = "I will come to school tomorrow"
q = "daddy will bring me to school"

Then p $\Leftrightarrow$ q = "I will come to school tomorrow IF AND ONLY IF daddy will bring me to school".

The truth value table for p $\Leftrightarrow$ q is given as:

P	q	p $\Leftrightarrow$ q
T	T	T
T	F	F
F	T	F
F	F	T

Note: The statement p $\Leftrightarrow$ q is only true when statements p and q are both true or both false.

3.5 Conditional Statements (or Implies)

This connective is usually denoted by the symbol $\Rightarrow$ placed between simple statements such as statement p $\Rightarrow$q read as “p IMPLIES q”.

Example

Let p = “EHJMC has won many awards of Excellence”
q = “Awards of excellence is meant for excellent performing schools”.

Then p $\Leftrightarrow$q = “If EHJMC has won awards of excellence THEN EHJMC is an excellent performing school. Another more simple form of the statement p $\Rightarrow$q is IF p THEN q.

The truth table for p $\Rightarrow$ q is given as:

P	q	p $\Rightarrow$ q
T	T	T
T	F	F
F	T	T
F	F	T

Note: The statement p $\Rightarrow$q is only false when the statement p is true and statement q is false.

3.6 Negation of Statements

The negation of a statement p denoted by the symbol ~ or’ i.e. ~p (or p^1) is the statement “Not p”.

Example

Let p = “Mathematics is simple”
Then ~p = “Mathematics is Not simple”
The truth table for ~p depends on the truth value of p

Thus, the truth value table is given as:

P	~p
T	F
F	T

3.7 Converse Statements

If p and q are statements then if $p \Rightarrow q$ is $q \Rightarrow p$.

Example

Now, $p \Rightarrow q$ = If x + 4 = 10 THEN x = 6
And $q \Rightarrow p$ = If " x = 6"
Now, $p \Rightarrow q$ = If x = 6 THEN x + 4 = 10

So since $p \Rightarrow q$ and $q \Rightarrow p$ then $q \Rightarrow p$ is the converse statement of $p \Rightarrow q$.

3.8 Inverse Statement

If p and q are statements and if $p \Rightarrow q$ then the inverse statement of $p \Rightarrow q$ is $\sim p \Rightarrow \sim q$.

Example

Let p = "x + 4 = 10" then ~p = "x + 4 $\neq$ 10"
and q = "x = 6" then ~q = "x $\neq$ 6"

Now since if $p \Rightarrow q$ and $\sim p \Rightarrow \sim q$ then $\sim p \Rightarrow \sim q$ is the inverse of $p \Rightarrow q$.

3.9 Contra-positive Statements

If p and q are statements and if $\sim p \Rightarrow \sim q$ then the contra-positive statement of $p \Rightarrow q$ is $\sim q \Rightarrow \sim p$.

Example

Let p = "x + 4 = 10" then ~p = "x + 4 $\neq$ 10"
and q= " x=6 then ~q$\neq$6"
now $\sim p \Rightarrow \sim q$ = " If x + 4 $\neq$10 Then x $\neq$ 6"
and $\sim q \Rightarrow \sim p$ = " If x + 4 $\neq$10 Then x $\neq$ 6"

3.10 Tautology and Contradiction

A compound statement which is always TRUE irrespective of the truth values of the simple statement is called a TAUTOLOGY.

A compound statement which is always FALSE irrespective of the truth values of the simple statements is called a CONTRADICTION.

Example

1) Determine which of the following are statements

i. Abuja is in Ghana
ii. 5+2=7
iii. Where is Sister Blessing?
iv. The Switch is on.
v. 3>5
vi. All square are parallelograms
vii. Do not over sleep tonight
viii. If 3-1=4 then an orange is yellow

Solutions

Statements i, ii, iv, v, vi and viii are statements while others are not.

Example

Use the truth table to prove the following statements

a) $p \vee (p \wedge q) = q$
b) $p \vee (q \vee r) = (p \vee q) \vee r$

Solutions

a) $p \vee (p \wedge q) = q$

P	q	$(p \wedge q)$	$P \vee (p \wedge q)$
T	T	T	T
T	F	F	T
F	T	F	F
F	F	F	F

Hence $P \vee (p \wedge q) = q$

c) $p \vee (q \vee r) = (p \vee q) \vee r$

P	q	r	(qvr)	(pvq)	P v(qvr)	(pvq) v r
T	T	T	T	T	T	T
T	T	F	T	T	T	T
T	F	T	T	T	T	T
T	F	F	F	T	T	T
F	T	T	T	T	T	T
F	T	F	T	T	T	T
F	F	T	T	F	T	T
F	F	F	F	F	F	F

a) Hence P v(qvr) = (pvq) v r

3.11 Laws of Algebra of Statements

Given the statement p, q, and r the following are the laws of algebraic logical statements.

1) The commutative laws
   a) pvq = qvp
   b) $p\wedge q = q \wedge p$
2) The Associative Laws
   a) pv(qvr) = (pvq)vr
   b) $p\wedge(q \wedge r) = (p\wedge q) \wedge r$
3) The Distributive Laws
   a) $p\wedge(qvr) = (p\wedge q)v(p\wedge r)$
   b) $pv(q\wedge r) = (p\,vq) \wedge (pv\,r)$
4) The De-Morgan's Laws
   a) $(pvq)^1 = P^1 \wedge q^1$
   b) $(p\wedge q)1 = p^1 \wedge q^1$
5) The Idempotent Laws
   a) pvp= p
   b) $p\wedge p = p$
6) Laws of Complementation
   a) $pvp^1 = T$
   b) $p\wedge p^1 = F$
   c) $(p^1) = p$

4.0 CONCLUSION

Logic is a philosophical theory of reasoning: the branch of philosophy that deals with the theory of deductive and inductive arguments and aims to distinguish good from bad reasoning system or instance of reasoning. However, logic has a mathematical approach to determining the truthfulness of a statement. The concept of Logic will help a social science researcher in drawing conclusions from varying information.

5.0 SUMMARY

In this unit, we have considered the following;

- a compound statement is a statement which consists of two simple statements or sub-statements
- the statement $p \vee q$ is false when both p and q are false, otherwise $p \vee q$ is true
- the statement $p \vee q$ is true when both p and q are true, otherwise $p \wedge q$ is false
- two compound statements are said to be logically equivalent if they have the same truth value.
- The truth table technique is used to establish whether or not logical statement are equivalent
- A compound statement which is always true irrespective of the truth value of the sub-statement is called Tautology.

6.0 TUTOR-MARKED ASSIGNMENT

Use the truth table to prove the following

1. $p \Leftrightarrow q = (p \Rightarrow q) \wedge (q \Rightarrow p)$
2. $(p \wedge q) \Rightarrow (pvq) is\ a\ autology$
3. $p \wedge \{(\sim p \wedge p) v (\sim p \wedge \sim q)\}\ is\ a\ contradiction.$

7.0 REFERENCES/FURTHER READING

Egbe, G. A. Odili and O. O Ugbebor (2005). *Further Mathematics*. Onitsha, Nigeria: Africana First Publishers Limited.

MODULE 2

UNIT 1 DIAGRAMMATIC PRESENTATION OF DATA

CONTENTS

1.0 INTRODUCTION

Having figures in form of raw data, it is always easy and convenient to understand when they are represented in diagrams. Since the aim of statistical methods is to present data in a more understandable way and render them intelligible, therefore, diagram brings about easy comparison and it should be noted that it should be used only where comparisons are called for since they give only approximate ideas.

2.0 OBJECTIVES

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

- present a raw data in form of diagram
- draw pictogram, bar char and pie chart
- distinguish pictogram from bar chart and pie chart
- understand which of the diagram will best explain the kind of the data available

- differentiate the different types of bar chart.

3.0 MAIN CONTENT

3.1 Pictogram

This entails the use of pictures to explain or illustrate statistical data. The pictures can be in form of; picture of men, cars, bottles, pots, houses etc. The pictures or drawings give quick and easy meaning to statistical data.

Examples

The population of a particular country from 1960 -1963 is represented on a diagram below:

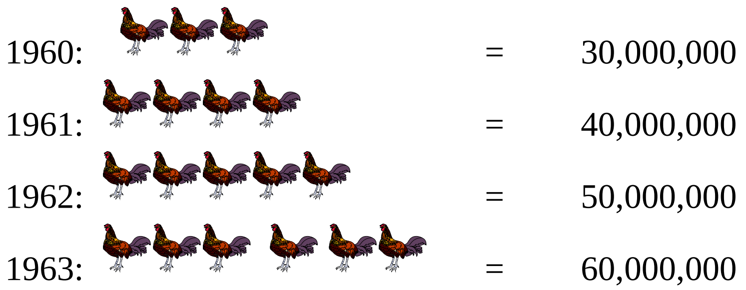


Fig. 1.1

The above diagram is used to illustrate how population of a country can be represented through pictures. From the above pictogram we can easily see that each picture represent 10,000,000 people.
Another example is shown below where each ball represents a student who gets the grade shown.

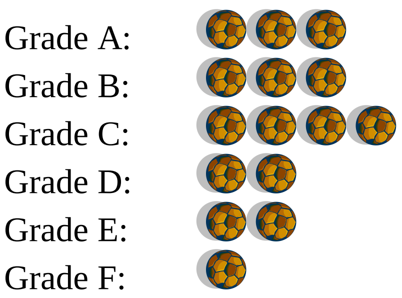


Fig. 1.2

3.2 Bar Chart

This is a type of diagram in which information given are represented with bars. The length of each bar is associated or corresponds to the value given and the width of each bar is equal to the other. There are three types of bar charts, which are Simple, Component and Multiple Bar charts. We will now look into each of them in details.

3.2.1 Simple Bar Chart

Only one information is represented on this type of bar chart as shown in the table below.

Example

The number of students taken for admission in various faculties of a particular University in 2007/2008 academic session is given below:

Faculties	Number of Students
Art	150
Agriculture	100
Business study	350
Engineering	100
Law	200
Science	300

Present the information on a bar chart?

Solution

Title of the chart: Bar Chart showing the number of students taken for admission.

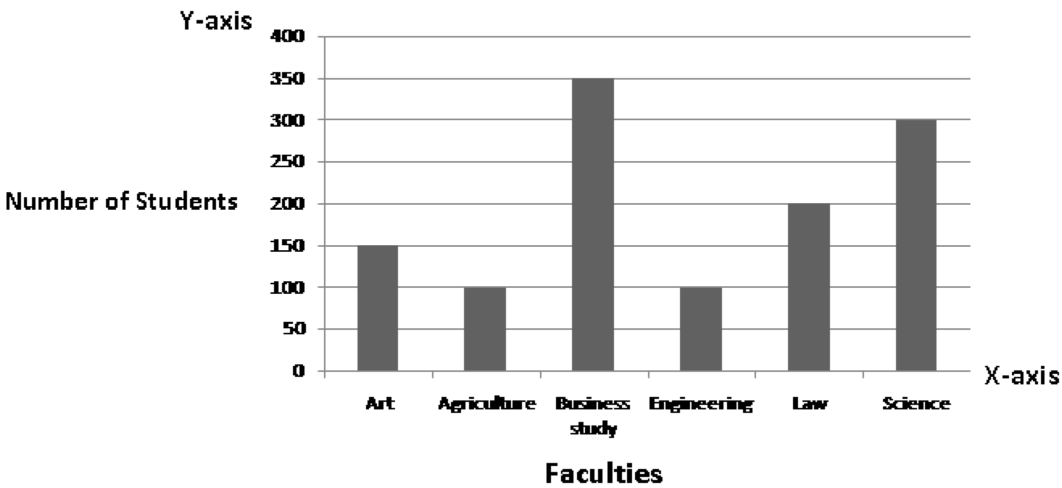


Fig. 1.3

The above is an example of a Simple Bar chart where faculties are represented on X-axis or horizontal axis and their corresponding number of students taken for admission is represented on the Y-axis (i.e. vertical axis).

3.2.2 Component Bar Chart

This exists when frequency can further be divided that is, there exist a sub-division within the frequency.

Example

Considering the example used for Simple Bar Chart, if the number of students taken for admission can further be classified by gender then we have:

Faculties	Male	Female
Art	100	50
Agric	60	40
Business Student	150	200
Engineering	80	20
Law	100	100
Science	150	150

Present the information on a bar chart.

Solution

Title of the chart: Component Bar Chart of the number of students taken for admission.

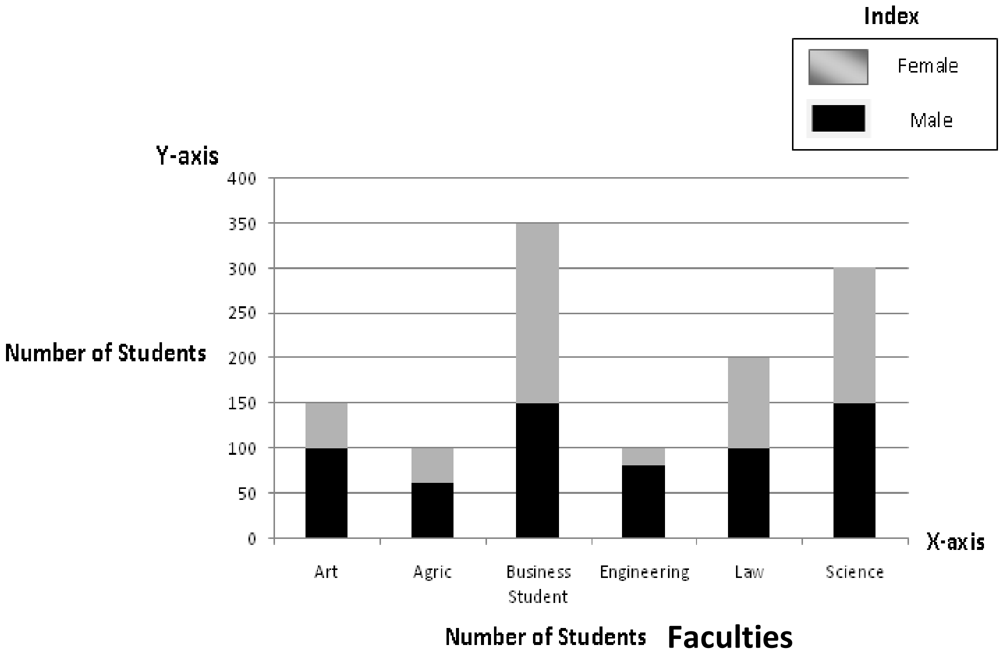


Fig. 1.4

The diagram in Fig 1.3 is a typical example of a component bar chart in which bar is being divided into two genders and female represents the upper part and the lower part of the bar is represented by the number of male given admission. There is always an index at the top right hand side of the graph to indicate the portion of each shaded area and to give a simple explanation of how each bar is being drawn.

3.2.3 Multiple Bar Charts

In a multiple bar chart, two or more set of inter-related data are represented. The method of simple bar chart is employed only that each bar is differentiated using different shades, colours etc.

Example

The information used for Component Bar chart is also employed to describe multiple bar charts.

Title of the chart: Multiple Bar Chart for the number of students taken for admission.

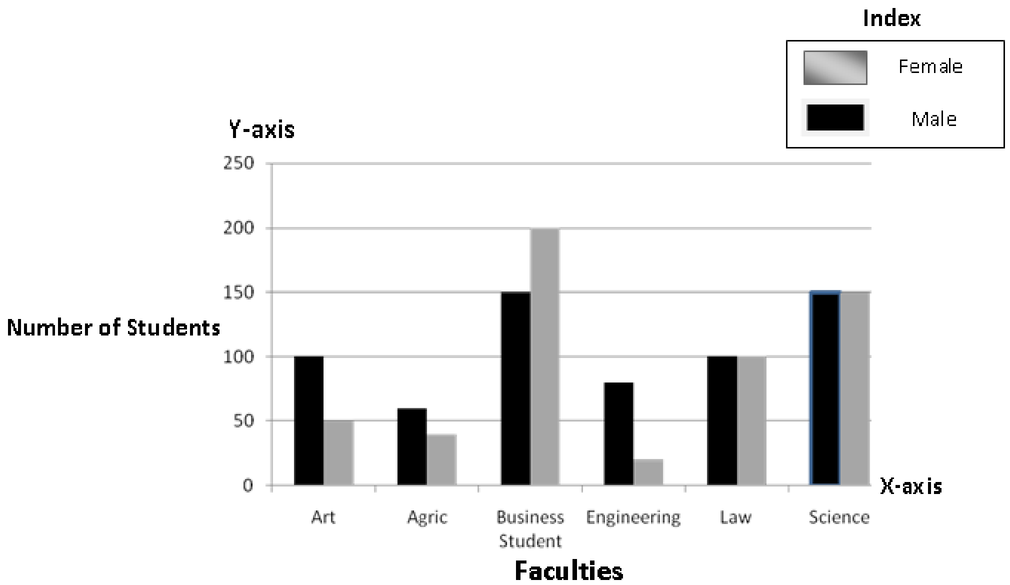


Fig. 1.5

This is an example of multiple bar charts.

3.3 Pie Chart

Another method of explaining data through diagram is pie chart. It is a chart in which each frequency is converted to degree and is presented on a circle, which is called pie chart. It is a graph in the shape of a circular pie.

Example

The number of passengers that board a bus in Iyana-Ipaja garage on a daily basis for a week is given below.

Days	**Passengers**
Monday	50
Tuesday	80
Wednesday	60
Thursday	60
Friday	150
Saturday	150
Sunday	50
Total	600

Calculation of degrees for each day.

Total passengers = 600

Monday: $\frac{50}{600} \times \frac{360}{1} = \frac{18000}{600} = 30^0$

Tuesday: $\frac{80}{600} \times \frac{360}{1} = \frac{28{,}800}{600} = 48^0$

Wednesday: $\frac{60}{600} \times \frac{360}{1} = \frac{21600}{600} = 36^0$

Thursday: $\frac{60}{600} \times \frac{360}{1} = \frac{21600}{600} = 36^0$

Friday : $\frac{150}{600} \times \frac{360}{1} = \frac{54000}{600} = 90^0$

Saturday: $\frac{150}{600} \times \frac{360}{1} = \frac{54000}{600} = 90^0$

Sunday: $\frac{50}{600} \times \frac{360}{1} = \frac{18000}{600} = 30^0$

Note

When you sum all the calculated values, it must be equal to 360^0, since the sum of all angles in a circle is 360^0. We will use the calculated values as shown below to plot the Pie Chart.

Monday	30^0
Tuesday	48^0
Wednesday	36^0
Thursday	36^0
Friday	90^0
Saturday	90^0
Sunday	30^0

Pie chart

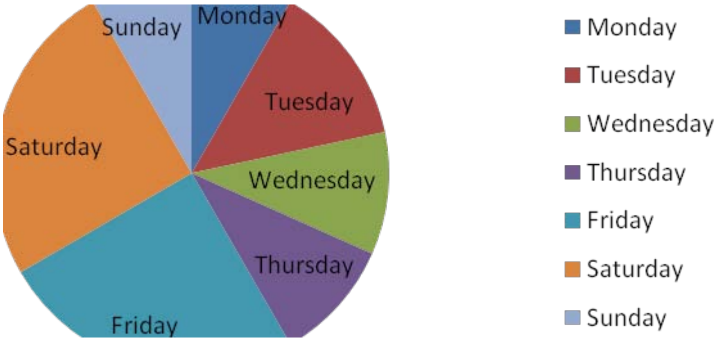


Fig 1.6

Fig 1.6 is the pie chart for the passengers that board a bus in Iyana-Ipaja for a particular week.

Example

We present fig.1.2 in a tabular form to plot a pie chart.

Grade	A	B	C	D	E	F
No of Students	3	3	4	2	2	1
Calculated Angles	72^0	72^0	96^0	48^0	48^0	24^0

Pie Chart

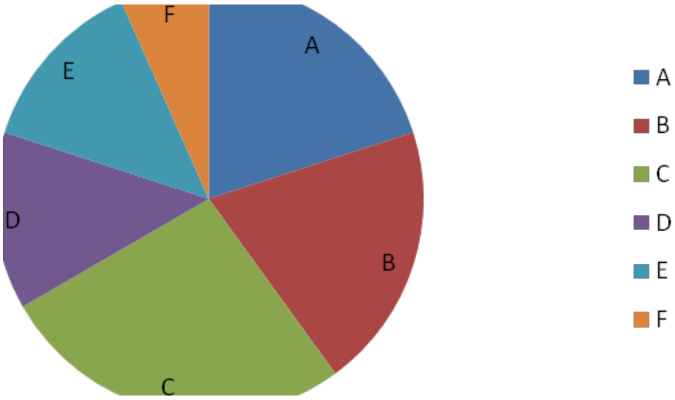


Fig 1.7

In fig 1.7, the size of each sector represents the number of students who get the grade shown in that sector.

4.0 CONCLUSION

The representation of statistical data in pictorial forms makes the application of statistics wider and gives easy interpretations to collected data. Especially where complex statistical data are presented in a simple diagram and charts which are true reflections of the situation described in the analysis, it simplifies the complexity of such data.

5.0 SUMMARY

This unit has helped you to understand the importance of diagram in statistical analysis. Various types of diagram have also been examined. The best way to come up with a good diagram was clearly explained.

6.0 TUTOR-MARKED ASSIGNMENT

1. Distinguish between various types of bar chart.
2. From the information in the table below, draw a suitable bar chart and a pie chart.

Year	Amount of Export
1990	60
1991	170
1992	150
1993	190
1994	220

1995	200

3. What is the importance of diagram in statistics?
4. A survey was carried out at a particular point in time to know the number of skilled and unskilled labour in some companies. The result is presented in the table below.

Name of Companies	Skilled labour	Unskilled labour
Rotsol Companies	120	80
Olawumi construction company	240	160
Wadet Printing Press	80	100
Markov Company	130	110
Solace Construction company	400	50

Draw a suitable diagram to explain the result of the survey.

5. Draw a pictogram for the information given in question (2).

7.0 REFERENCES/FURTHER READING

Gupta, C. B. & Vijcy Gupta (23rd Revised Edition 2004). *An introduction to Statistical Methods*. New Delhi: Vikas Publishing House PVT Ltd.

Allan G. Bluman, (2004). *Elementary Statistics: A Step by Step Approach*. New York: McGraw-Hill Companies, Inc.

UNIT 2 FREQUENCY DISTRIBUTION

CONTENTS

1.0 INTRODUCTION

After a researcher might have gotten a raw data from any source, there is a need for the raw data to be arranged and organised in a meaningful way in order to be able to describe and come up with a useful inference. The method that is being used for such organisation and arrangement is called frequency distribution.
Frequency means the number of times something happens. For example, four students got Grade B. The frequency of grade A is four. Frequency distribution simply means a tabular arrangement of data into the class they belong to.

2.0 OBJECTIVES

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

- differentiate between grouped and ungrouped frequency distribution
- draw a frequency table set of raw data
- identify and define terms that are associated with frequency distribution.

3.0 MAIN CONTENT

3.1 Importance of Frequency Distribution

The most important form of tabulation is the frequency distribution. When dealing with a large form of quantitative data such as the weight of three hundred athletes, it is convenient to group them into classes; the obese, average weight and underweight could form the different classes. The total number of athletes that are obese is the frequency of that class, the total number of athletes that are of average weight is called the frequency of the averagely weighted and the total number of athletes that are underweight is also the frequency of the underweight.

3.2 Ungrouped Frequency Distribution

This is a type of frequency distribution in which data are not compressed together in a particular interval.

Example

Given a set of raw data below, construct a frequency table for it.

2, 3, 2, 1, 4, 4, 1, 1, 3, 2, 5, 3, 5, 1, 5, 1, 4, 2, 3, 0,
4, 1, 0, 3, 5, 2, 5, 0, 4, 1, 4, 0, 5, 2, 3, 0, 4, 4, 1, 2.

Solution

Table 3.1

Ungrouped Frequency Distribution

Value	Tally	Frequency
0	~~IIII~~	5
1	~~IIII~~III	8
2	~~IIII~~II	7
3	~~IIII~~I	6
4	~~IIII~~III	8
5	~~IIII~~I	6

Table 3.1 is an example of how ungrouped frequency distribution should be. It is obtained by counting the number of 0, 1, 2, 3, 4, 5 in the raw data and recording the number of times they occur in front of it as the frequency. In order to avoid double counting or mistake in counting, the method of crossing the one counted by a dash may be adopted.

SELF ASSESSMENT EXERCISE 1

The numbers 2,2,2,3,4,2,5,3,6,3,4,2,6,4,2, can be represented in a tabular form for easy interpretation as shown below.

Table 3.2

Item	Frequency
2	6
3	3
4	3
5	1
6	2

In Table 3.2, the number of times a particular item occurs is called the frequency in that distribution.

SELF ASSESSMENT EXERCISE 2

Tabulate the height of members of your tutorial group as shown in tables 3.1 and 3.2 above.

3.3 Grouped Frequency Distribution

Many times, we involve ourselves in a large volume of data that have close numerical values. The purpose of classification is to organise the data to a manageable size. To achieve this, data is grouped. This grouping may be of equal or unequal interval. This further reduces the task involved in analysis, when the number of observation becomes larger. Observation is thus grouped into a number of classes. In doing this, we have to decide at the beginning the number of groups or classes we wish to classify the data into. Each of the group is given as an interval and it is called class interval.

Example

The following are scores obtained by forty students who sat for POL 221 examination in National Open University of Nigeria. Construct a frequency distribution for the scores.

56	20	45	70	50	49	62	39	41	65
25	76	59	48	55	57	71	49	42	44
63	60	40	45	50	31	35	21	58	56
54	56	63	30	39	28	49	53	64	66

Solution: Table 3.3

Grouped Frequency Distribution

Scores	Tally	Frequency
20-29	IIII	4
30-39	~~IIII~~	5
40-49	~~IIII~~ ~~IIII~~	10
50-59	~~IIII~~ ~~IIII~~I	11
60-69	~~IIII~~II	7
70-79	III	3

The table above shows how a grouped frequency distribution should look like. The scores obtained by students are compared together in an interval of 20-29, 30-39...70-79 and the corresponding number of students was recorded as frequency.

SELF ASSESSMENT EXERCISE 3

Consider the frequency table showing the population of a country and use it to answer the questions below.

Age Group	No. of persons in Millions
0- 9	49
10-19	39
20-29	31
40-49	27
50-59	18
60-66	14
70-79	10
80 and above	2

a. Determine the population of the country above.
b. Which age group has the lowest population in millions?
c. Which age group has the highest population?

Answers are obvious from the table above. Can you try it?

3.4 Other Related Concepts

3.4.1 Class Interval

Class interval is a set of classes that are used to define the raw data or size of the group chosen. It can be determined by finding the range of the raw data obtained and dividing it by the value of number of classes we desire to have.

Using table 3.2 of grouped frequency distribution, class interval is regarded as the scores which are: 20-29, 30-39................70-79.

3.4.2 Class Limit

These are the end numbers of class interval. The lower value for class interval is called lower class limit while the upper value for class interval is called upper class limit.

Table 3.2 can also be used to explain this. For class interval 20-29, 20 is the lower class limit and 29 is the upper class limit. For class interval 30-39, 30 is the lower class limit and 39 is upper class limit etc.

3.4.3 Class Boundaries

Class boundaries are easily gotten by subtracting 0.5 from lower class limit or lower value of class interval and adding 0.5 to upper class limit or upper value of class interval. For example, class boundaries for Table 3.2 are 19.5 - 29.5, 29.5 - 39.5 ....69.5 - 79.5

3.4.4 Class Mark

This is the mid-point or value of the class interval. It can be derived by adding lower and upper class limit and dividing by two (2) or adding lower and upper class boundaries and dividing the sum by two (2).
Referring to table 3.2 that we have been using, class mark for the first class interval is

$$\frac{20+29}{2} = \frac{49}{2} = 24.5 \; Or \; \frac{19.5+29.5}{2} = 24.5$$

3.4.5 Class Width

This is the size of the class interval and it is obtained by subtracting lower class boundaries from upper class boundary.

Using table 3.2 again, class width for the first class 19.5 - 29.5 is
$29.5 - 19.5 = 10$

The second class boundary is
$39.5 - 29.5 = 10$
Therefore class width for class interval of table 3.2 is 10.

4.0 CONCLUSION

To describe situations, draw conclusions, or make inferences about events, the researcher must organise the data in some meaningful way. The most convenient way of organising data is to construct a frequency distribution.

5.0 SUMMARY

The importance of frequency distribution in statistical analysis cannot be over emphasised, therefore we have been able to explain the types of frequency distribution and some terms used in forming frequency distribution.

6.0 TUTOR-MARKED ASSIGNMENT

1. Differentiate between grouped and ungrouped frequency distribution.
2. The number of cars that passed through University of Ibadan main gate at the interval of five minutes was recorded for a particular day as follows;
   4, 3, 3, 4, 9, 2, 0, 6, 8, 5, 3, 2, 4, 1, 0, 5, 4, 3, 6, 8, 10, 11, 8, 7, 12, 0, 6, 8, 10, 7, 5, 12, 11, 10, 6, 7, 7, 9 1, 8, 10, 1, 10, 7, 5, 0, 6, 9, 3, 2,
   Construct a suitable frequency distribution for the data?
3. A frequency distribution table for weekly wages (in Naira) of workers in a particular company in Nigeria is given below.

Weekly Wages	Frequency
50-54	5
55-59	4
60-64	9
65-69	6
70-74	8
75-79	10
80-84	8
85-89	3

Find:

(a) Class boundaries for the table
(b) Class width
(c) Class mark for the 4th class
(d) Class limit for the 6th class.

7.0 REFERENCES/FURTHER READING

Amazing, J.C (ed.). (1991). *Introductory Mathematics* 1: *Algebra. Trigonometry and Complex Numbers*. Onitsha: Africana/Fep Publishers Ltd.

Bluman, A. G. (2004). *Elementary Statistics: A Step by Step Approach*. New York: McGraw-Hill Companies, Inc.

UNIT 3 GRAPHICAL PRESENTATION OF DATA

CONTENTS

1.0 INTRODUCTION

One of the best ways of explaining statistical data is through graphs. Graphs are used to present statistical data in a simpler way. Instead of having bar alone for each class, a graph uses continuous scales in rectangular coordinates. Better inferences can also be made with graphs than just diagram. In this unit, we will be looking at various types of graphs that can be used to explain statistical data.

2.0 OBJECTIVES

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

- identify various types of graph
- draw a histogram to represent a set of data
- draw frequency polygon from frequency distribution
- draw a cumulative frequency table
- draw a cumulative frequency curve.

3.0 MAIN CONTENT

3.1 Histogram

Histogram is a graph in which class boundaries or class interval is marked on the horizontal axis and the corresponding class frequency on the vertical axis.

Note that the bars of a histogram must be joined together and this differentiates it from bar chart.

Example

Table 3.1

Considering the frequency distribution table below, draw a histogram for the table.

Class boundaries	Frequency
0.5-5.5	6
5.5-10.5	8
10.5-15.5	5
15.5-20.5	4
20.5-25.5	2

Solution

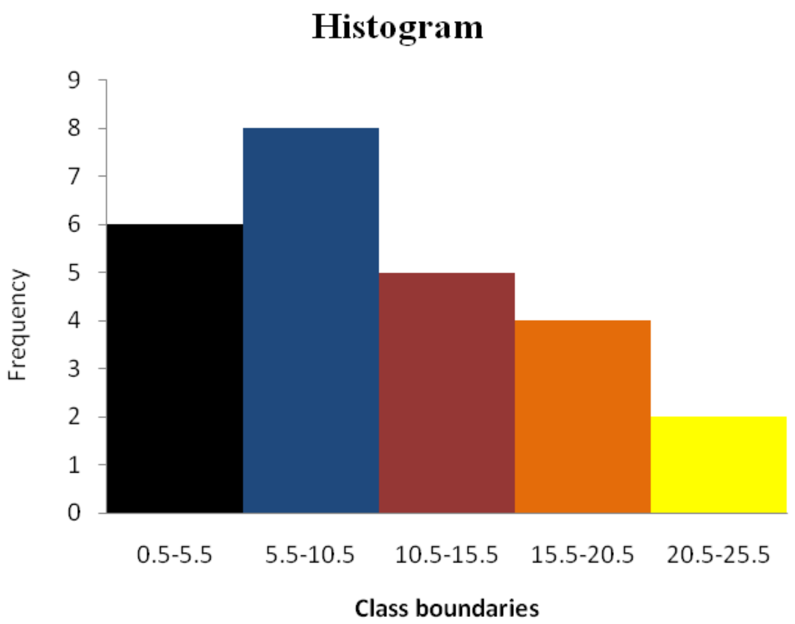


Fig.3.1

Figure 3.2 is an example of a typical histogram where we have class boundaries against corresponding frequency.

3.2 Frequency Polygon

Frequency Polygons are very useful if we want to compare two distributions. It is very easy to draw two or more polygons on the same graph; this may not be possible in the case of histogram. The only setback to this kind of comparison is that the total frequency must be the same, in order to give a fair comparison.

This is a graph that joins the mid-point of the tops of column of the histogram through straight lines. To draw a frequency polygon, there is need to obtain class marks for each class boundary against the corresponding frequency.

Example

The following data were obtained from the result of a test a teacher conducted for his students.

68	65	50	40	46	56	53	70	60	85
45	46	90	95	70	86	88	69	94	53
56	57	69	85	46	93	73	75	54	51
43	72	65	64	68	54	75	77	69	43
52	56	60	75	63	45	80	86	78	66

Construct:
(a) Frequency distribution.
(b) Histogram and frequency polygon
(c) Frequency polygon.

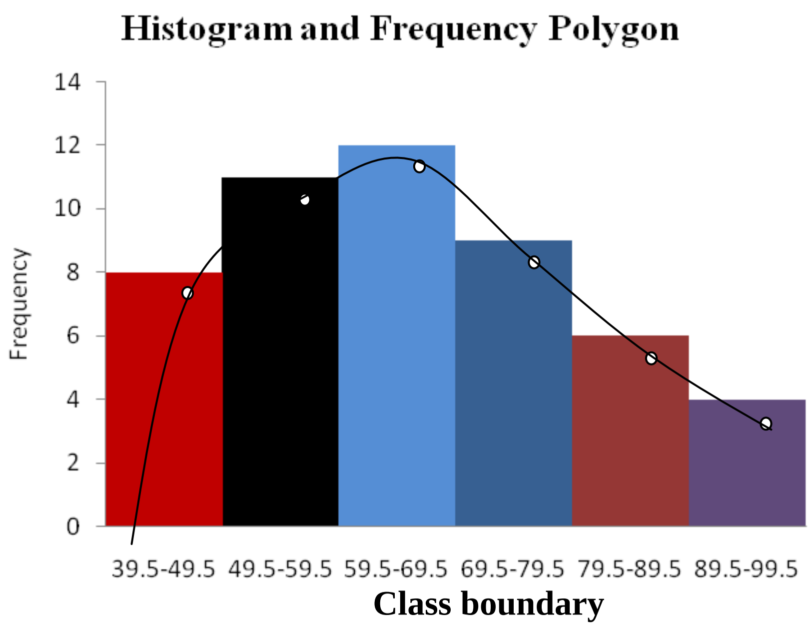
Solution

(a) Frequency Distribution

Table 3.2

Class interval	Class boundaries	Class mark	Frequency
40-49	39.5-49.5	44.5	8
50-59	49.5-59.5	54.5	11
60-69	59.5-69.5	64.5	12
70-79	69.5-79.5	74.5	9
80-89	79.5-89.5	84.5	6
90-99	89.5-99.5	94.5	4

(b) Histogram and Frequency Polygon



(c) Frequency Polygon

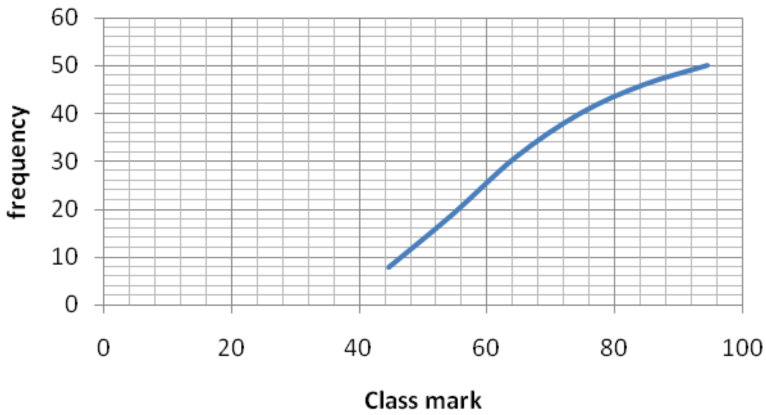


Fig. 3.2

3.3 Cumulative Frequency Curve

Cumulative frequency curve is also called Ogive graph. Before this curve can be plotted, cumulative frequency distribution is needed and it can be obtained by adding the values of frequency together then cumulative frequency is plotted against class marks to give cumulative frequency curve.

Example

Using the question given under polygon, construct a cumulative frequency distribution and cumulative frequency curve.

Class boundaries	Class mark (x)	Frequency (f)	Cumulative Frequency (cf)
39.5-49.5	44.5	8	8
49.5-59.5	54.5	11	19
59.5-69.5	64.5	12	31
69.5-79.5	74.5	9	40
79.5-89.5	84.5	6	46
89.5-99.5	94.5	4	50

Cumulative Frequency Curve

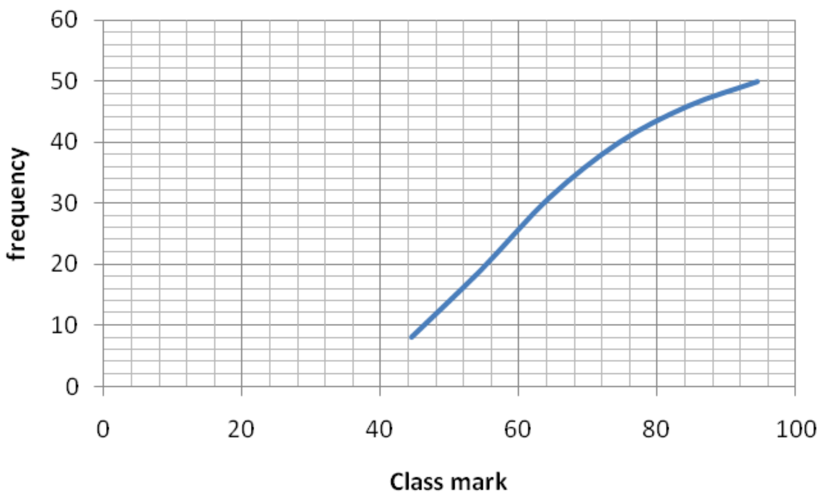


Fig. 3.3

4.0 CONCLUSION

From the discussions above, it shows clearly that statistical data can be represented in graphical forms, which can help researchers to carry out further research on the sample collected. In a broad sense, graph is a diagram that shows relationships between numbers. Graphs arrange numerical information into a picture from which it is often possible to see overall patterns or trends in the information.

5.0 SUMMARY

This unit has dealt with using graph to represent statistical data. Various types of graphs have been considered and distinctions among them were clearly explained.

6.0 TUTOR - MARKED ASSIGNMENT

1. What is the importance of graphs in statistical data?
2. Explain the various types of graph that we have.
3. A set of data was obtained and recorded below.

10	10	15	25	26	13	37	24	19	45
46	12	19	36	38	18	26	48	14	24
35	33	20	48	34	33	29	16	39	46

Draw: (a) A Histogram
(b) A Frequency Polygon and
(c) An Ogive graph

7.0 REFERENCES/FURTHER READING

Bluman, A. G. (2004). *Elementary Statistics- A Step by Step Approach.* New York: McGraw- Hill Companies, Inc., 1221 Avenue of the Americas.

Gupta, C. G. and Gupta, V. (2004). *An Introduction to Statistical Methods*. New Delhi: Vikas Publishing House Pvt Ltd.

Microsoft ® Encarta ® 2009 © 1993-2008 Microsoft Corporation**.**

UNIT 4 MEASURES OF CENTRAL TENDENCY

CONTENTS

1.0 INTRODUCTION

Very often, when we are given a set of numerical data, we may want to look for a single quantity, which represents the entire sets. Thus, this may at times, make us to disregard the entire members of the set. A statistical measure, which describes the middle or centre of a set of data, is called measure of central tendency. In this unit, we will explore the common measures of central tendency, which are; the Arithmetic Mean, Mode and Median.

2.0 OBJECTIVES

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

- calculate the Arithmetic Mean
- obtain the median of group and ungrouped distribution
- calculate the mode.

3.0 MAIN CONTENT

3.1 The Arithmetic Mean

The arithmetic mean (or just mean) is the most important measure of central tendency, the reason being that, all members of the set are used in the calculation of the mean. It is however affected by the extreme values of the set unlike the range, which will be discussed in the next unit under measure of dispersion.

Let

$$X = x_1 + x_2 + x_3 + x_4 + \cdots x_n$$

Then the mean usually denoted as $\bar{x}$ is given as:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 \ldots + x_n}{n}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Note:

The symbol $\sum$ is called sigma, which is used in mathematics to denote **Summation**.

Example

Find the mean of the following set of numbers: 10, 9, 11, 13, 12, 12, 11, 13, 10 and 16

Solution:

Mean

$$\bar{x} = \frac{10 + 9 + 11 + 13 + 12 + 12 + 11 + 13 + 10 + 16}{10}$$

$$\bar{x} = \frac{117}{10} = 11.7$$

If the distributions: x_1, x_2, x_3... x_n have frequencies f_1, f_2, f_3, ...,f_n respectively then:

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \ldots + f_{nx_n}}{f_1 + f_2 + f_3 + f_4 + \cdots f_n}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum f_i}$$

Example

Find the mean of the set of data in the table below:

Mark x_i	0	1	2	3	4	5	6	7	8	9
Frequency f_1	2	3	4	6	1	4	2	2	1	3

Solution

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum f_i}$$

$$= \frac{(0 \times 2) + (1 \times 3) + (2 \times 4) + (3 \times 6) + (4 \times 1) + (5 \times 4) + (6 \times 2) + (7 \times 2) + (8 \times 1) + (9 \times 3)}{2 + 3 + 4 + 6 + 1 + 4 + 2 + 2 + 1 + 3}$$

$$= \frac{114}{28}$$

$$\bar{x} = 4.07$$

3.1.1 Using Assumed Mean

We use the above example to calculate the mean using the assumed mean.

Let $x_1 = d_2 + A$
$x_2 = d_2 + A$
$x_3 = d_2 + A$
$x_i = d_i + A$

Then $\sum_{i=1}^{n} x_1 = \sum_{i=1}^{n} d_1 + NA$

$$\Longrightarrow \frac{\sum_{i=1}^{n} x_1}{N} = \frac{\sum_{i=1}^{n} d_1}{N} + \frac{NA}{N}$$

$$\bar{x} = \bar{d} + A$$

The constant A is called the ASSUMED or GUESSED MEAN. d_i is the deviation of x_i from the assumed mean.

So the $\bar{x} = A + \frac{\sum_{i=1}^{n} f_i d_i}{\sum_{i=1}^{n} f_i}$

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

Example

A farmer recorded the mass of 25 timbers as follows:

10	14	12	10	12	11	11	9	13
16	13	9	12	13	12	10	15	
10	9	11	8	14	12	8	11	

a) Construct a frequency table for the data.
b) Use an assumed mean of 12 kg to calculate the mean.

Solution:

Given assumed mean, A= 12 kg

Masses (x)	Frequency (f)	d =x – A	*Fd*
8	2	-4	-8
9	3	-3	-9
10	4	-2	-8
11	4	-1	-4
12	5	0	0
13	3	1	3
14	2	2	4
15	1	3	3
16	1	4	4
	$\sum f = 25$		$\sum fd = -15$

Using the formula:

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

$$= 12 + \frac{(-15)}{25}$$

= 11.4 kg

3.1.2 Arithmetic Mean of Group Data

In our earlier discussion of grouped frequency distribution, it was mentioned that the values between any class interval are considered as condensed at the mid- point of the class interval or class mark. If x_i is the class mark of the i^{th} class interval then the mean $\bar{x}$ of the grouped frequency distribution is defined as:

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

Example

The following table shows the distribution of weekly wages earned by 60 casual employees of INEC in Lagos State of Nigeria.

Wages	No. 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 (A) of ₦ 74.5, calculate the mean of the distribution.

Solution

Class interval	Class centre (x)	Frequency(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

		$\sum f = 60$		$\sum fd - 270$

So

$$\bar{x} = A + \frac{\sum_{i=1}^{n} fd}{\sum f}$$

$$74.5 + \frac{(-270)}{60}$$

$$\bar{x} = ₦\, 70.00$$

3.2 The Median

When data consisting of "n" members are arranged in order of magnitude, the middle values or member is called THE MEDIAN of the data if "n" is odd and when "n" is even the mean of the middle values is the MEDIAN. Generally, the two median of a set of "n" members is defined as the $1/2\ (n+1)^{th}$ value when the "n" members are arranged in order of magnitude.

3.2.1 Median of an ungrouped Data

When dealing with ungrouped data, to calculate the median is very simple. All we need to do is to rearrange the data in the order of magnitude and simply bring out the median. If the data is an even number, then we find the mean of the two middle values.

Example

Find the Median of: 3, 2, 2, 5, 1, 4, 3, 2, 1, 5 and 2

Solution

By arranging in ascending order: 1, 1, 2, 2, 2, 2, 3, 3, 4, 5, 5

The Median in the above data is 2

Example

Find the Median of: 104, 107, 110, 103, 118, 100

Solution

By re-arranging: 100, 103, 104, 107, 110, 118

$$\text{median} = \frac{104 + 107}{2} = 105.5$$

3.2.2 Median of a Grouped Data

For a grouped frequency distribution, the class interval which contains the median value is called the Modal Class.

Example

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

Mass (kg)	Frequency
15 – 24	4
25 – 29	35
30 – 34	49
35 – 39	24
40 – 49	6
50 – 60	2

Find the median.

Solution

Since the total frequency is 120

The median is at the $\left(\frac{60+61}{2}\right)^{th}$ class

Hence median class = (30 – 34)

3.3 The Mode

The mode of a distribution is the value, which occurs most frequently in the distribution. It is the value, which has the highest frequency in the distribution. In the grouped frequency distribution, the class interval with the highest frequency is called Modal Class.

3.3.1 Mode of an Ungrouped Distribution

Example

The marks of 40 students out of 10 marks in Mathematics test are as follows:

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

(a) Draw a frequency table for the distribution
(b) State the mode and median of the distribution
(c) Calculate the mean of the distribution

Solution: Construct a frequency Distribution table

Mark	Frequency
1	8
2	7
3	3
4	4
5	3
6	4
7	3
8	2
9	3
10	3

(a) Mode = 1 mark
(b) Median = $\frac{4+4}{2}$ = 4 marks
(c) Mean = $\frac{(1\times8)+(2\times7)+(3\times3)+(4\times4)+(5\times3)+(6\times4)+(7\times3)+(8\times2)+(9\times3)+(10\times3)}{40}$

Mean = $\frac{180}{40} = 4.5$

3.4 Median from Grouped Data through Histogram

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

(a) From a Histogram
(b) From a cumulative frequency curve.

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. Equally, the value on the variable axis which divides the area of the histogram in the ratio 1:3 is called the FIRST QUARTILE. While the value on the variable axis which divides the areas of the histogram in the ration 3:1 is called the THIRD QUARTILE. Notably, the median coincides with the SECOND QUARTILE. The 1^{st}, 2^{nd}, 3^{rd} quartiles are usually denoted by Q_1, Q_2, and Q_3 respectively.

Also the area of the Histogram can be divided into 100% so that the value on the variable axis which divides the area of the Histogram in the ration 1:99 is called the FIRST PERCENTILE.

$\Rightarrow$ Median coincides with 50^{th} percentile
$\Rightarrow$ Q_1 coincide with 25^{th} percentile
$\Rightarrow$ Q_2 coincide with 75^{th} percentile

These are usually denoted as P_{25}, P_{50}, P_{75}. These will be discussed in detail in the subsequent units.

Example

The following table shows the distribution of marks (in percentages) scored by a class of forty students in a promotion examination.

Marks	No. of Students
10 – 29	6
30 – 39	5
40 – 49	7
50 – 59	10
60 – 69	5
70 – 79	4
80 – 89	3

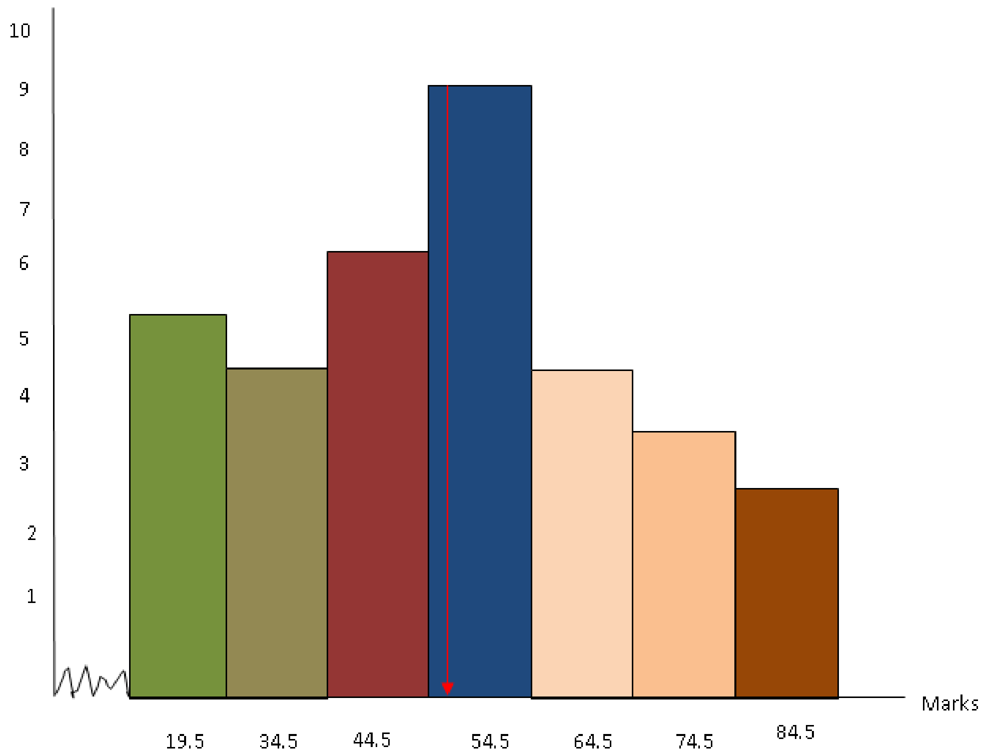
a. Calculate the mean marks
b. Draw a histogram of the distribution and use it to estimate the median mark

Solution

Mass (kg)	**Class centre (C)**	**Frequency (f)**	**C x f**	**Class boundary**
10 – 29	19.5	6	117	9.5 – 29.5
30 – 39	34.5	5	172.5	29.5 – 39.5
40 – 49	44.5	7	311.5	39.5 – 49.5
50 – 59	54.5	10	545.5	49.5 – 59.5
60 – 69	64.5	5	322.5	59.5 – 69.5
70 – 79	74.5	4	298.5	69.5 – 79.5
80 – 89	84.5	3	253.5	79.5 – 89.5

		$\sum f = 40$	$\sum = 2020$	

(a) Mean $\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum f_i} = \frac{2020}{40} = 50.5$



The median value is obtained from the histogram by drawing vertical line, which divides the histogram into two equal parts**.**

3.5 Mode from grouped data through histogram

In a grouped frequency distribution, the class with the highest frequency is called the MODAL CLASS. When data are given raw, it may not be easy to determine the mode especially when it is large. The easiest way to determine the mode is to classify the data forming a frequency table for it. From the frequency table it is very easy to determine the mode or modal class respectively. The mode can also be determined geometrically from a histogram of a grouped frequency distribution.

Example:

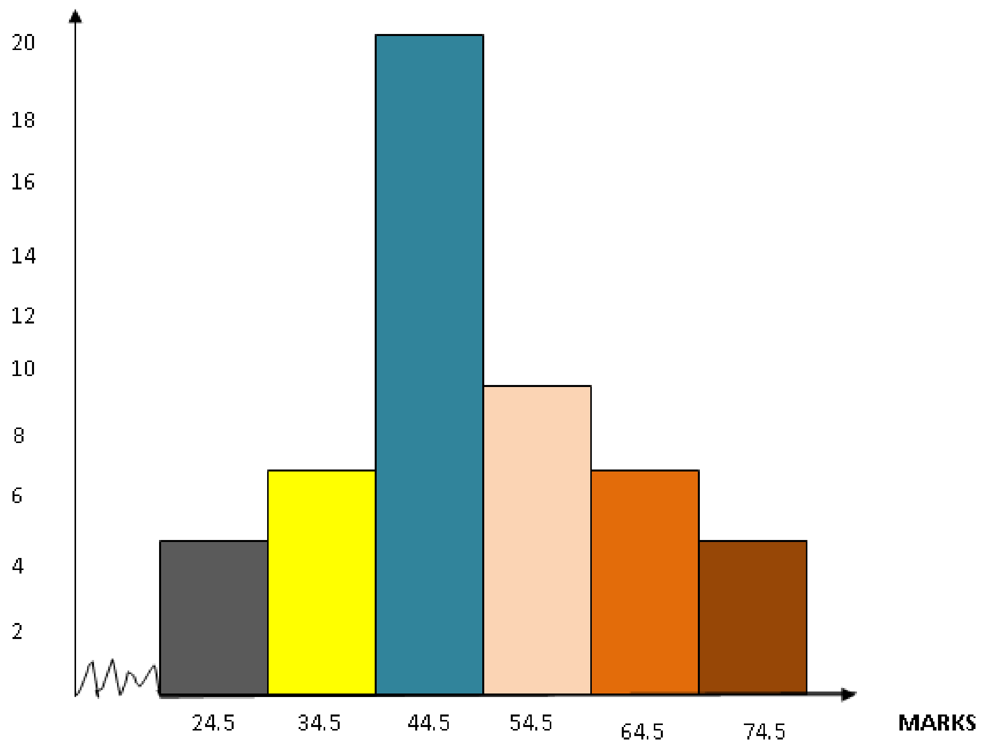
The table below gives the distribution of marks of 50 students in an examination.

Marks (%)	20 – 29	30 – 39	40 – 49	50 – 59	60 – 69	70 - 79

No. of Students	4	7	20	9	6	4

(a) Draw a Histogram for the distribution.
(b) Use your histogram to estimate, correct to one decimal place, the mode.

Marks (%)	**Class Mark**	**Frequency**	**Class boundary**
20 – 29	24.5	4	19.5 – 29.5
30 – 39	34.5	7	29.5 – 39.5
40 – 49	35.5	20	39.5 – 49.5
50 – 59	45.5	9	49.5 – 59.5
60 – 69	54.5	6	59.5 – 69.5
70 - 79	64.5	4	69.5 – 79.5



Hence the mode is 44.5

4.0 CONCLUSION

The mean possesses special properties that make it the most frequently used. However, the sensitivity of the mean of extreme scores values that are not balanced on both sides of the distribution makes the median the measure of choice when distributions are marked uneven or skewed. The median is also the measure of choice when there are open-ended classes to which specific values cannot be assigned.

5.0 SUMMARY

We demonstrated the calculation of the three central tendencies that are frequently used for description of the central features of frequency distributions and arrays of scores; the mean, the mode and the median.

6.0 TUTOR - MARKED ASSIGNMENT

1. The table below shows the description of marks (in percentages) scored by a class of forty students in a promotion examination. Calculate the mean mark.

Mass (kg)	Frequency
10 – 29	6
30 – 39	5
40 – 49	7
50 – 59	10
60 – 69	5
70 – 79	4
80 – 89	3

2. The table below gives the distribution of marks of one hundred candidates in a Mathematics examination.

Mark	1-20	21-30	31-40	41-50	51-60	61-70	71-85
frequency	10	14	17	22	19	9	9

(a) Draw a histogram of mode.
(b) Draw a cumulative frequency curve and use it to estimate:

(i) The mode
(ii) The median

3. The masses of 37 students are shown in the table below.

Masses (kg)	40	41	42	43	44	45
No. of Students	5	10	6	4	7	5

(a) State the mode and median of the distribution.
(b) Calculate the mean of the distribution.

4. In an objective test marked out of 40, the marks scored by 35 students out of 40 students are given in the table below:

Marks (%)	1 – 5	6 – 10	11 - 20	21 – 30	31 - 35	36 - 40
No. of Students	2	7	12	8	5	1

a) Calculate the mean of the distribution

Draw a histogram and use it to estimate

b) The median
c) The mode of the distribution

7.0 REFERENCES/FURTHER READING

Richard, P. Runyon and Audrey, Haber (1996). *Fundamentals of Behavioural Statistics*, (8th ed.). USA: McGraw-Hill Companies, Inc.

Walker and McLean, (1973). *Ordinary Statistics.* Edward Arnold Publisher Ltd.

Robert D. Mason; Douglas A. Lind & William G. Marshal, (1999). *Statistics Techniques in Business and Economics*. USA: McGraw Hill Companies, Inc.

UNIT 5 MEASURES OF PARTITION 1

CONTENTS

1.0 INTRODUCTION

There are different ways, depending on the purpose of partitioning data array. Partitioning allows us to have relative observation of one element in a set of data to others. The middle value (or arithmetic mean of the two middle values) of a set of data arranged in order of magnitude that divides (partitions) the set into two equal parts is the median. Though median is also a measure of partition, most texts limit it to measure of central tendency. In addition, some of these texts refer to measure of partition as measure of position, location quantity or relative standing. Partitioning set of data into two equal parts can be extended to those values that partition the set into four equal parts, ten equal parts and hundred equal parts. These are called quartiles, deciles, and percentiles respectively.

2.0 OBJECTIVES

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

- define and use these partitioning measures; quartiles, deciles, percentiles, to partition any given data array
- identify the position of a data value in a data set, using measures of partition; percentiles, quartiles, deciles, deciles, e.t.c.
- make a relative observation of one element in a set of data to others using percentiles, quartiles and deciles.

3.0 MAIN CONTENT

3.1 Quartiles

Quartile partitions the array into four equal groups denoted by Q_1, Q_2, and Q_3. The lower quartile (first quartiles) is Q_1, the middle quartile (second quartile) is Q_2 which is also the median of any distribution and the upper quartile (third quartile) is Q_3.

3.1.1 Computation of Quartiles

These quartiles are computed as follows:

Step 1: Arrange the data in order of magnitude from the lowest

Step 2: Locate the middle of the data by finding the median of the data. This is the value for Q_2 (second quartile)

Step 3: The median of the data below Q_2 is the lower quartile, Q_1

Step 4: The median of the data above Q_2 is the upper quartile, Q_3

Example

Find Q_1 Q_2 and Q_3 for the data set 16, 25, 14, 18, 21, 23, 19, 18, 21, 49

Solution:

Step 1: Arrange the data in order
14, 16, 18, 18, 19, 21, 21, 23, 25, 49

Step 2: Find the Median (Q_2)
The two values at the middle are 19 and 21
Then Q_2 $\frac{19+21}{2} = 20$

Step 3: The data below Q_2;
14, 16, 18, 18, 19
The value at the middle is 18
Then $Q_1 = 18$

Step 4: The data above Q_2
21, 21,23,25,49
The value at the middle is 23
Then $Q_3 = 23$

3.2 Deciles

Deciles position the data array into ten equal groups, separated by D_1, D_2, D_3… D_9. The fifth deciles, D_5 are the Median.

3.2.1 Computation of Deciles

The deciles corresponding to a given value, (x) is computed using the formula;

$$Decile = \frac{(number\ of\ values\ below\ x) + 0.5}{total\ number\ of\ values}\ X\ 10$$

A data value (D V) corresponding to a given deciles is computed by $DV = \frac{n\ X\ d}{10}$

Step 1: Arrange the given data in order of magnitude

Step 2: Substitute into the formula
Where: n = total number of values
d = deciles given

Step 3: When D V is a decimal value, round it off, then count from the lowest value of the data to the corresponding round off DV.

Step 4: When D V is a whole value, count from the lowest value of the data up to the mid value of D Vth and (D V+1)th values.

Example

In a test administered to 10 students, the total score is 30. Find the deciles rank of a score of 26.

18, 28, 25, 26, 24, 30, 22, 20, 16, 27

Solution

Arrange the data in ascending order.
16, 18, 20, 22, 24, 25, 26, 27, 28, 30

The formula is

$$Decile = \frac{(number\ of\ values\ below\ x) + 0.5}{total\ number\ of\ values}\ X\ 10$$

Number of values below 26 = 6, therefore the solution is

$$Decile = \frac{6+0.5}{10}\ X\ 10 = 6.5$$

That is, a student with score 26 performed better than $\frac{6.5}{10}$ of the class.

Example

Using the data given above, find the value data that corresponds to the 8^{th} Deciles.

Step 1: Arrange the data in ascending order
16, 18, 20, 22, 24, 25, 26, 27, 28, 30

Step 2: Substitute into the formula $DV = \frac{n \; X \; d}{10}$

$$DV = \frac{10 \; X \; 8}{10} = 8$$

Step 3: The DV is 8, (Which is a whole number). Therefore, the mid value of the 8^{th} and 9^{th} will be the corresponding 8^{th} deciles.

That is, $\frac{27+28}{2} = 27.5$

3.3 Percentiles

A Percentile is another measure of partition that partitions data array into hundred equal groups. Percentiles are used in educational and health-related matters to indicate the position of an individual in a group and most often used for large data sets. Example of its use is to compare an individual's test score with the national norm. For instance, in a Joint Admissions Matriculation Board (JAMB) test, a student's score was 248 and this placed the student at the 70^{th} percentile in the data of scores. Where will the student's score be partitioned in relation to the scores of others in the test? The answer is obvious since the student's score was partitioned at the 70^{th} percentile, this means 70% of all the test scores were lower than the student's score and 30% were higher. Though percentiles are similar in sound as percentages, yet they are not the same. For example, if a student scores 84 marks out of 100, the student has a percentage of 84, and not necessarily 84 percentile, since this score might be the highest, lowest, or middle score.

3.3.1 Computation of Percentiles

The percentile corresponding to a given value x is computed by:

$$Percentile = \frac{(number\ of\ values\ below\ x) + 0.5}{total\ number\ of\ values} \; X \; 100\%$$

While a data value corresponding to a given percentile is computed by:

$$DV = \frac{n\,X\,p}{100}$$

Step 1: Arrange the given data in order of magnitude.
Step 2: Substitute into the formula.
Step 3: When data value (DV) is a decimal value, round it off, then count from the lowest value of the data up to the corresponding round off DV.
Step 4: When DV is a whole value, count from the lowest value of the data up to the mid-value of D Vth and (D V + 1)th values.

Example

A Mathematics teacher administered a test to 20 students and the total mark obtainable was 40. These are the scores of students who took the test;

22, 20, 18, 25, 28, 32, 38, 22, 26, 34,
18, 22, 22, 30, 35, 33, 25, 27, 33, 27

Find the percentile rank of a score 28.

Solution

Arrange the given data in the following ascending order of magnitude.
18, 18, 20, 22, 22, 22, 24, 25, 25, 26, 27, 27, 28, 30, 32, 33, 33, 34, 33, 38

$$Percentile = \frac{(number\ of\ values\ below\ x) + 0.5}{total\ number\ of\ values}\ X\ 100\%$$

$$= \frac{12+0.5}{20}\ X\ 100\%$$

= 62.5th percentile
$\cong$ 63th percentile

This implies that a student with the score 28, performed better than 62.5% of the class.

Example

Determine (a) The 25th and (b) the 60th percentiles for the data in the above example

Solution

Arrange the given data in the following ascending order.

18, 18, 20, 22, 22, 22, 24, 25, 25, 26, 27, 27, 28, 30, 32, 33, 33, 34, 33, 38

(a) Counting $\left(\frac{25N}{100}\right)^{th}$ where N is the total number of values in the data from the lowest.

$P_{25} = \left(\frac{25 X 20}{100}\right)^{th}$ from the lowest value.

= 5^{th} from the lowest value

= 22

(b) $P_{60} = \left(\frac{60N}{100}\right)^{th}$

$= \left(\frac{60 X 20}{100}\right)^{th}$

= 12^{th}

= 27

3.4 Computation for Grouped Data

For grouped distribution, we will make use of the formula below to calculate the quartiles, Deciles and Percentiles. Since quartiles divides a set of distribution into four equal portions, the first and the third quartile Q_1 and Q_3 will be $1\left(\frac{n+1}{4}\right)^{th}$ and $3\left(\frac{n+1}{4}\right)^{th}$ items respectively in a distribution.

And to find the 8^{th} Deciles = $8\left(\frac{n+1}{10}\right)^{th}$ item,

Also, $P_{34} = 34\left(\frac{n+1}{100}\right)^{th}$

In general, if a particular quartile splits a distribution into k equal parts, the jth quartile of the distribution is $j\left(\frac{n+1}{K}\right)^{th}$ item of the size ordered distribution.

Example

Consider the distribution below.

Class interval (x)	Frequency (F)	Cumulative Frequency
70-72	5	5
73-75	18	23
67-78	42	65
79-81	27	92
82-84	8	100

From above: $n = \sum f = 100$

The 1st quartile Q_1 is given by $\left(\frac{100+1}{4}\right)^{th}$ and

The 3rd quartile Q_3 is given by $3\left(\frac{100+1}{4}\right)^{th}$ items

It implies that Q_1 is the 25.25th item and Q_3 is the 75.75th item.

Therefore, since Q_1 occurs in the class 76-78 hence it is the Q_1 class and also since Q_3 occurs in the 79-81, hence it is the 3rd quartile Q_3 class.

The general formula for Q_1, Median (Q_2), and Q_3 quartiles are as follow:

$$Q_1 = l_1 + \left(\frac{\frac{N+1}{4}(\sum f)_1}{fQ_1}\right)c_1$$

$$Median = Q_2 = l_1 + \left(\frac{\frac{N+1}{2}(\sum f)_{med}}{fmed}\right)c$$

$$Q_3 = l_3 + \left(\frac{3\left(\frac{N+1}{4}\right)(\sum f)_3}{fQ_1}\right)c_3$$

Where

l_1 and l_3 are the lower class of the boundaries of the 1st and 3rd quartiles classes,
N is the total number of items in the distribution
$\sum f_{1\ and}\ \sum f_3$ are the respective cumulative frequencies lower than the respective quartiles classes.$fQ_1\ and\ fQ_3$
$(\sum f)_{med}$ sum of the frequencies of all classes lower than median class
"C" is the class width
$l_1 in\ Median\ =$
$Q_2\ is\ the\ lower\ class\ boundary\ of\ the\ median\ class$
$fmed$ = frequency of median class

Note that, the above formula is applicable to both deciles and percentiles as well, but for deciles, 4 is replaced with 10, and for percentiles, 4 is replaced with 100 respectively.

Examples

Find using an interpolation formula method to calculate the median, quartiles and the 37th percentile of the weight of 1200 Chickens given below.

Weight in (gms)	Frequency (f)	Cumulative Frequency (cf)
56-58	73	7
59-61	1	20
62-64	68	88
65-67	144	232
68-70	197	429
71-73	204	633
74-76	208	841
77-79	160	1001
80-82	101	1102
83-85	54	1156
86-88	25	1181
89-91	13	1194
92-94	4	1198
95-97	2	1200

Solution

The cumulative is calculated as shown in the table above.

a) Median

$$Q_2 = l_1 + \left(\frac{\frac{N+1}{2}(\sum f)_{med}}{fmed}\right)c$$

$=\frac{120+1}{2}$= 600.5th item

Median class is 71-73

l_1=70.5

$\sum f = 429$

$frequency\ of\ median\ class = 204$

$c = 3$

$$median = 70.5 + \left(\frac{600.5-429}{204}\right)3$$

$$= 73.02$$

b) $Q_1$is the $\frac{1200+1}{4}$th item =$300.25th\ item$

Q_1 Class $68 - 70$

$fQ_1 = 197, c = 3$

$$Q_1 = 67.5 + 3\left(\frac{300.25 - 232}{197}\right) = 68.54$$

$$Q_3 \frac{3(1200 + 1)}{4} = 900.75th\ item$$

$l_3 = 76.5(\sum f)_3 = 841,$

$fQ_3 = 160, C = 3$

$$Q_3 = 76.5 + 3\left(\frac{900.75 - 841}{160}\right) = 77.62$$

c) To find the 37th percentile

$$P_{37} is\ the\ 37\left(\frac{1200 + 1}{100}\right) th\ item\ 444.37\ item$$

This lies in the class 71-73

$l_{37} = 70.5 (\sum f)_{37} = 429,$
$fP_{37} = 204$, C=3

$$P_{37} = 70.5 + 3\left(\frac{444.37 - 429}{200}\right) = 70.72$$

4.0 CONCLUSION

We have been dealing with qualitative aspect of a distribution before now; we have however been able to look at some aspects of a distribution, which can be described in quantitative terms by calculating certain values from it.

5.0 SUMMARY

Measures of partition are ways of grouping/partitioning distribution data to have relative partition of each value in the data. These measures treated are Quartiles, Deciles and Percentiles. Quartiles, Deciles and Percentiles partition distribution data into 4 (four), 10 (ten), and 100 (hundred) respectively.

6.0 TUTOR - MARKED ASSIGNMENT

1. What do you understand by the following?

a. Quartile
b. Deciles
c. Percentiles

2. The table below represents the data for weights of boys in a school. Find the approximate weights corresponding to each percentile given;

a) 25^{th} b) 40^{th} c) 50^{th} d) 69^{th}

Weight (kg)	Frequency
42.5 – 45.5	9
45.5 – 48.5	12
48.5 – 51.5	17
51.5 – 54.5	22
54.5– 57.5	15

3. (a) Is it possible to express all quartiles as percentiles? Support your answer with example.

(b) Is it possible to express all deciles as percentiles? Support your answer with example.

7.0 REFERENCES/FURTHER READING

Richard, P. Runyon & Audrey, Haber (1996). *Fundamentals of Behavioural Statistics*, (8th ed.). USA: McGraw-Hill Companies, Inc.

Walker and McLean (1973). *Ordinary Statistics*. Edward Arnold Publisher Ltd.

Robert D. Mason; Douglas A. Lind & William G. Marchal, (1999). *Statistics Techniques in Business and Economics.* USA: McGraw Hill Companies Inc.

Microsoft ® Encarta ® 2009. © 1993-2008 Microsoft Corporation**.**

UNIT 6 MEASURES OF DISPERSION II

CONTENTS

1.0 INTRODUCTION

In data description, statisticians go beyond measures of central tendency and measures of partition. They also go much further to include measures of dispersion. Measures of central tendency describe the mid-points of data array through different ways: measures of partition/position/location/ relative standing or quartiles describe the partitioning of data array into groups while measures of dispersion/variation describe the spread-out of data values from each other.

2.0 OBJECTIVES

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

- understand and define measures of dispersion
- give and define its measures
- calculate these measures from any given data set.

3.0 MAIN CONTENT

3.1 Dispersion

Consider the table below of weekly wages of ten employees in different companies.

Weekly Wages	Employees in Company A	Employees in Company B	Employees in Company C	Employees in Company D
₦ 5, 000.00	4	10	2	0
₦ 10, 000.00	4	0	4	8
₦ 15, 000.00	4	0	8	4
₦ 20, 000.00	4	0	4	8
₦ 25, 000.00	4	10	2	0

The mean wage of each company is ₦ 15, 000.00, but the way in which the wages are spread out about the centre (mean) is different from each company. A measure of dispersion is a statistic signifying the extent of the spread-out. For the spread of a data set, some of these measures of dispersion shall be treated in this unit. These measures to be considered are:

- Range,
- Semi-inter-quartile Range, or simply Quartile Deviation,
- Inter-percentile Range, and
- Mean Deviation.

3.2 Range

The easiest but crudest way of measuring the spread/scatter of a data set is the range of values. This is defined as the difference between the largest and the smallest values of a data set. In a frequency distribution, the range is taken to be the difference between the lower limit of the class at the lower extreme of the distribution, and the upper limit of the class at the upper extreme.

Example

The range of the set of numbers;
8, 4, 10, 12, 6, 7, 8, 15, 14, is
15 – 4 = 11.

The highest value in the range is 15, while the lowest value is 4. Therefore, the lowest value is subtracted from the highest value, giving the range. From the above example, we can see that the greater the range the greater the dispersion. Range is easy to understand, and its computation is simple, but is not based on all observations. Also, range doesn’t change if values of a set of data change provided the two extreme values remain the same.

3.3 Semi-Inter-Quartile Range or Quartile Deviation

This is another measure of dispersion which is better than the range. Quartiles, as earlier defined partition data array into four equal parts. The lower quartile, Q_1 is the first ¼ of the values of the data array. The middle quartile, Q_2 (usually called the median), is the first ½ of the values of the data array. The upper quartile, Q_3 is the first ¾ of the values of the data array. The fourth quartile, Q_4, is the whole data array. The inter-quartile range is $Q_3 – Q_1$, and the semi inter-quartile range is ½ $(Q_3 – Q_2)$. The inter-quartile range is 50% of the values in a data set or dividing this into two, would be equal to the difference between the median and any quartile of a data array.

Example

Find the;

a. Range
b. Semi-inter-quartile range for the table below

Mark	**Class mark**	**Frequency (f)**	**Cumulative frequency (cf)**
60 – 62	61	5	5
63 – 65	64	18	23
66 – 68	67	42	65
69 – 71	70	21	92
72 – 74	73	8	100

Solution

a. The lower limit of the lower class is 59.5
The upper limit of the upper class is 74.5
The Range is 74.5 – 59.5 = 15

b. The lower quartile Q_1, =

$$Q_1 = l_1 + \left(\frac{\frac{N+1}{4} - (\sum f)_1}{fQ_1} \right) c_1$$

$$Q_2 = l_1 + \left(\frac{\frac{N+1}{2} - (\sum f)_{med}}{fmed} \right) c$$

$$Q_3 = l_3 + \left(\frac{3\left(\frac{N+1}{4}\right) - (\sum f)_3}{fQ_1} \right) c_3$$

$$= 65.5 + \frac{\frac{1x100}{4} - 23}{42} \times 3$$

$$= 65.5 + \frac{25 - 23}{42} \times 3$$

= 65.5+0.14

= 65.64

The upper quartile

$$Q_3 = 68.5 + \frac{\frac{3 \times 100}{4} - 64}{27} \times 3$$

$$= 68.5 + \frac{75 - 65}{27} \times 3$$

$$= 68.5 + 1.11$$

$$= 69.11$$

Semi-inter-quartile range

$$= \frac{1}{2}(Q_3 - Q_1)$$

$$= \frac{1}{2}(69.11 - 65.64)$$

$$= \frac{1}{2}(3.47) \ = 1.74$$

Example

1. Find the range of the frequency distribution given below.

Mark	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45	46-50
Frequency	0	2	4	11	20	10	5	3	2	1

Solution

The lower class =1-5
Lower limit of the class = 0.5
The upper class = 46-50
Upper lower limit of the upper class = 50.5

Therefore, the range = 50.5-0.5
= 50

2. Find the range of the set 12, 6, 7, 3, 15, 10, 18, 5
Solution
Range = largest number – smallest number
= 18-3 =15

3.4 Quartile Deviation

1. Find the quartile deviation of the example above

Solution

Lower quartile,

$$Q_1 = \left(\frac{\frac{N+1}{4} - (\sum f)}{f}\right)c$$

$$= 15.5 + \frac{\frac{61}{4} - 6}{11} \times 5$$

$$= 15.5 + 4.2$$

$$= 19.7$$

Upper quartile,

$$Q_3 = Q_2 = l_3 + \left(\frac{\frac{N+1}{4}(\sum f)_{med}}{fmed}\right)c$$

$$= 25.5 + \frac{45.75 - 37}{10} \times 5$$

$$Q_3 = 25.5 + 4.38$$

$$= 29.88$$

Quartile deviation

$$= \frac{1}{2}(Q_3 - Q_3)$$

$$= \frac{1}{2}(29.88 - 19.7)$$

$$= 5.09$$

The size of a quartile deviation gives an indication about the uniformity of the values of a data set. That is, if the quartile deviation is small it means large uniformity.
Quartile deviation is a measure of partition and not a measure of dispersion since it does not show the spread of values around an average, but shows distance on scale. Although quartile deviation is simple to compute and easy to understand, it is not based on all observations and may not be affected by little fluctuations of sampling.

3.5 Mean Deviation

This is a much better measure of dispersion. Mean deviation takes into consideration the limitations of range and quartile deviation; it takes account of all observations. Thus, as it implies,it is the mean of the deviations or differences of the scores from either the mean, median or mode. In other words, mean deviation is the mean of the absolute values of the deviation from some measure of central tendency. The most commonly used of the measure of central tendency for computing mean deviation is the median followed by the mean. This is the average deviation from median.

The following examples illustrate the styles involved in the calculation of mean deviation.

Examples

140 students sat for Mathematics test, their marks are as shown in the frequency distribution table below. Calculate the mean deviation of their marks.

Marks	**Class marks (x)**	**Frequency (f)**	**Cumulative Frequency (f)**	**Absolute Deviation from median |d| = $\|X - median\|$**	**f|d|**
0-10	5.5	4	4	57.5	230
11-20	15.5	6	10	47.5	285
21-30	25.5	9	19	37.5	337.5
31-40	35.5	2	28	27.5	247.5
41-50	45.5	12	40	17.5	210
51-60	55.5	20	60	7.5	150
61-70	65.5	42	102	2.5	105
71-80	75.5	22	124	12.5	275
81-90	85.5	10	134	22.5	225

91-100	95.5	6	140	32.5	195
		140			$\sum f\lvert d\rvert = 2260$

Median = the size of $\frac{140}{2}th$ value

$$= l + \frac{\frac{N+1}{2} - \sum f}{f} \times c$$

Where l= lower limit of the median class
c= median class width
f = frequency of the median class
$\sum f$ = cumulative frequency below the median class

$$\text{median} = 60.5 + \frac{\frac{141}{2} - 60}{42} \times 10$$

$$= 60.5 + \frac{10.5}{42} \times 10$$

$$= 63.0$$

Then, mean deviation

$$= \frac{\sum f\lvert d\rvert}{\sum f}$$
$$= \frac{2260}{140}$$
$$= 16.143$$

Example

Find the range of the frequency distribution given below.

Mark	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45	46-50
Frequency	0	2	4	11	20	10	5	3	2	1

Solution

The lower class =1-5
Lower limit of the class = 0.5
The upper class = 46-50
Upper lower limit of the upper class = 50.5
Therefore, the range = 50.5-0.5

= 50

2 Find the range of the set 12, 6, 7, 3, 15, 10, 18, 5

Solution

Range = largest number – smallest number

= 18-3

=15

Quartile deviation

3 Find the quartile deviation of the example above

Solution

Lower quartile, $Q_1 = l + \left(\frac{\frac{N+1}{4} - \sum f}{f_c}\right)c$

$= 15.5 + \frac{\frac{61}{4} - 6}{11} \times 5$

$= 15.5 + 4.2$

$= 19.7$

Upper quartile, $Q_3 = l + \left(\frac{\frac{3(N+1)}{4} - \sum f}{f}\right)c$

$= 25.5 + \frac{45.75 - 37}{10} \times 5$

$Q_3 = 25.5 + 4.38$

$= 29.88$

Quartile deviation $= \frac{1}{2}(Q_3 - Q_3)$

$= \frac{1}{2}(29.88 - 19.7)$

$= 5.09$

3.6 Standard Deviation

Standard deviation is an improvement of mean deviation. Two extra stages are added to the calculation of mean deviation; we obtain the most desirable measure of dispersion called standard deviation. This is the most satisfactory and universally adopted measure of dispersion. The measure of central tendency used in calculating standard deviation is mean. The example below illustrates the computation stages of standard deviation.

Steps

The steps involved in the calculation of standard deviation are as follow,

1. Arrange the data set in order (if not arranged)

2. Compute the frequency of each value (if not given)
3. Compute the mean, $\bar{x} = \frac{\sum fx}{\sum f}$
4. Find the deviation, $d = x - \bar{x}$ of each value from the mean, $\bar{x}$
5. Take the squares of the deviations (this eliminates the negative signs)
6. Compute the sum of the squares of the deviations, $\sum(x - \bar{x})^2$
7. Then, divide your result in 6 by $\sum f$, $i.e$ $\sum(x - \bar{x})^2 / f$
8. Finally take the square root of your result in 7.

Example

Calculate the standard deviation from the frequency distribution given blow.

X	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80
F	18	16	15	12	10	5	2	2

X	Mid point (X)	Frequency(F)	FX	Deviation x from mean d= $x - \dot{x}$	d^2	Fd^2
0-10	5.5	18	99.0	-21.625	467.641	8417.538
11-20	15.5	16	248.0	-11.625	135.141	2162.256
21-30	25.5	15	382.5	-1.625	2.641	39.615
31-40	35.5	12	426.0	8.375	70.141	841.692
41-50	45.5	10	455.0	18.375	337.641	3376.410
51-60	55.5	5	277.5	28.375	805.641	4625.705
61-70	655.0	2	131.0	38.375	2340.141	2945.282
71-80	75.5	2	151	48.375		4680.282
		80	**2170**			**2708.78**

Mean, $\bar{X} = \frac{\sum fx}{\sum f}$

$= \frac{2170}{80}$

=27.125

Then, standard deviation, S = $\sqrt{\frac{\sum fd^2}{\sum fx}}$

$= \sqrt{\frac{27088.78}{80}}$

$= \sqrt{338.610}$

= 18.401

3.6.1 Calculation of the Standard Deviation from an Assumed Mean

From the example given above, it can be seen that the mean is not a whole number. In such a case there is a truncation error and excessively heavy arithmetical calculation involved. To overcome theses difficulties we introduce the calculation of standard deviation using assumed mean.

The only difference from the mean calculation is that, a guess of an assumed mean is made.

Example

The frequency table below shows the distribution of marks in an examination.

Use an assumed mean to calculate the standard deviation of marks.

Mark (X)	1-5	6-10	11-15	16-20	21-25	25-30	31-35	36-40	41-45	46-52
Frequency (f)	0	2	4	11	20	10	5	3	2	1

Solution

Mark	Midpoint (x)	Frequency (f)	fx	Assume mean M=23 d=X – m	fd	d^2	fd^2
1-5	3	0	0	-20	0	400	0
6-10	8	2	16	-15	-30	225	450
11-15	13	4	52	-10	-40	100	400
16-20	18	11	198	-5	-55	25	270
21-25	23	20	460	0	0	0	0
26-30	28	10	280	5	50	25	250
31-35	33	5	165	10	50	100	500
36-40	38	3	114	15	45	225	675

41-45	43	2	86	20	40	400	800
45-50	48	1	48	25	25	625	625
		58	**1419**		**85**		**3975**

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

$= \sqrt{\frac{3975}{58} - \left(\frac{85}{58}\right)^2}$

= 8.1896

$\simeq 8.2\ (d.p)$

The above equation formula is for grouped data/frequency distribution. The following is for ungrouped data.

$$= \sqrt{\frac{\sum d^2}{N} - \left(\frac{\sum d}{N}\right)^2}$$

Example

From the data given below, calculate

a. Mean

b. i, Standard deviation using calculated mean

ii, Standard deviation using assumed mean

30, 33, 24, 28, 20, 17, 25, 39, 34, 42

a. Mean $\bar{X} = \frac{\sum x}{N} = \quad \bar{X} = 29.2$

X	$\boldsymbol{d = X - \bar{X}}$	$\boldsymbol{d^2}$
17	-12.2	148.84
20	-9.2	84.64
24	-5.2	27.04
25	-4.2	17.64
28	-1.2	1.44
30	0.8	0.64
33	3.8	14.44
34	4.8	23.04
39	9.8	96.04
42	12.8	163.84
$\sum X = 292$	0	577.6

Standard deviation, S = $\sqrt{\frac{\sum d^2}{N}}$

$= \sqrt{\frac{577.6}{10}}$

$= \sqrt{57.76}$

= 7.6

b. i.

X	**d=X-M, M= 28**	d^2
17	-11	121
20	-8	64
24	-4	16
25	-3	9
28	0	0
30	2	4
33	5	25
34	6	36
39	11	121
42	14	196
	12	592

Standard Deviation

SD = $\sqrt{\frac{\sum d^2}{N} - \left(\frac{\sum d}{N}\right)^2}$

$= \sqrt{\frac{592}{10} - \left(\frac{12}{10}\right)^2}$

$= \sqrt{59.2 - 1.44}$

= 7.6

Example

These are the heights of 30 girls. Calculate the mean and standard deviation of the girls' heights.

Height	148	146	150	152	154	156	158	160	162	164	166	168	170
Frequency	1	2	2	3	3	7	2	5	1	2	1	0	1

Solution

Height (cm) (X)	**Frequency (f)**	**f X**	$= X - \bar{X}$	d^2	fd^2
146	1	146	-10.47	+109.62	109.62
148	2	296	-8.47	+71.74	143.48
150	2	300	-6.47	+41.86	83.72
152	3	456	-4.47	+19.98	59.94
154	3	462	-2.47	+6.10	18.30
156	7	1092	-0.47	+0.22	1.54
158	2	316	1.53	2.34	4.68
160	5	800	3.53	12.46	62.30
162	1	162	5.53	30.58	30.58
164	2	328	7.53	56.70	113.40
166	1	166	9.53	90.82	90.82
168	0	0	11.53	132.94	0
170	1	170	13.53	183.06	183.06
2054	30	4694			901.44

Mean, $\bar{X} = \frac{\sum fx}{\sum f}$

$= \frac{4694}{30}$

$= 156.47$

Standard deviation, S = $\sqrt{\frac{\sum fd^2}{\sum f}}$

$= \sqrt{\frac{901.44}{30}}$

$= \pm\ 30.05$

$= \pm 5.48$

The solution of the above problem can be approached through assumed mean as follows:

Mean = Assumed mean + $\sqrt{\frac{\sum fd}{\sum f}}$

And standard deviation= $\sqrt{\frac{\sum fd^2}{\sum f} - (\frac{\sum fd}{\sum f})^2}$

Height (cm) (X)	Frequency (f)	Assumed =158 d =X-M	d^2	fd	fd^2
146	1	-12	144	-12	144
148	2	-10	100	-20	200
150	2	-8	64	-16	128
152	3	-6	36	-18	108
154	3	-4	16	-12	48
156	7	-2	4	-14	28
158	2	0	0	0	0
160	5	2	4	10	20
162	1	4	16	4	16
164	2	6	36	12	72
166	1	8	64	8	64
168	0	10	100	0	0
170	1	12	144	12	144
2054	30	1		-46	972

Mean = Assumed mean + $\frac{\sum fd}{\sum f}$

$= 158 + \frac{-46}{30}$

$= 158 + 1.53$

$= 156.47$

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

$= \sqrt{\frac{972}{30} - \left(\frac{-46}{30}\right)^2}$

$= \sqrt{32.4 + 2.35}$

$= \sqrt{30.05}$

± 5.48

SELF ASSESSMENT EXERCISE

1. Calculate the:

a) Range
b) Quartile deviation, and
c) Mean deviation from the following data:

x	**f**
0-10	18
11-20	16
31-40	12
41-50	10
51-60	5
61-70	2
71-80	2

2. Compute the

a) Range
b) Quartile deviation, and
c) Mean deviation from the following data
24, 14, 12, 16, 18, 20, 26, 18,
14, 28, 20, 22, 24, 16, 18, 22.

3. The table below gives the distribution of heights of 100 youths at an interview.

Height (cm)	156-160	161-163	164-166	167-169	170-175
Frequency	16	14	18	21	13

Using an Assumed Mean of 168cm, calculate correct to one decimal place:

a. The mean height
b. The standard deviation of the distribution.

4. The table below shows the marks obtained by one hundred and sixty candidates in a test.

Marks	**No. of Candidates**
0-09	18
10-19	28
20-29	37
30-39	40
40-49	16
50-69	21

Using an assumed mean of 34.5 or otherwise, calculate the mean to the nearest whole number

a. The Mean
b. The standard deviation of the distribution.

4.0 CONCLUSION

All the methods of measuring dispersion so far discussed are not universally adopted for want of adequacy and accuracy. The range is limited by the size of the difference between the two extreme values. Close look at quartile shows that, it is a partition rather than measure of dispersion. Mean deviation is fairly better out of the three measures of dispersion discussed but not a satisfactory method. During the computation processes there is no mathematical justification that deals with the negative deviations. These limitations called for the statisticians to seek for the best measure, a measure that is free from these limitations. These other measures of dispersions were also discussed in this unit.

5.0 SUMMARY

Measures of dispersion/variation describe the spread of values of a data set. Some of the measures discussed in this unit are range, semi-inter-quartile range also known as quartile deviation, mean deviation, and the generally accepted measure of dispersion which is the standard deviation, was also discussed in this unit. Examples were solved on each of these measures of dispersion, which should help the students to understand the concept been discussed and also aid to attempt the entire tutor-marked assignments.

6.0 TUTOR-MARKED ASSIGNMENT

1. Give the definition of the following:

a. Range of a data set
b. Inter-quartile range
c. Semi-inter-quartile range/quartile deviation
d. Mean deviation.

2. Use the following data sets below to discuss a range as a measure of dispersion.

a. 24, 14, 12, 16, 18, 20, 26, 14, 28, 20, 22,
b. 22, 20, 14, 24, 14, 26, 24, 10, 38, 20, 16,
c. 24, 24, 12, 24, 24, 24, 24, 24, 28, 24, 24.

3. Which would you consider best; Range, Quartile deviation or mean deviation, as a measure of dispersion. Support your conclusion with example(s).

7.0 REFERENCES/FURTHER READING

Allan, G. Bluman (2004). *Elementary Statistics. A Step by Step Approach*. New York: McGraw-Hill Companies.

Gupta, C. B. & Vijcy Gupta (2004). *An Introduction to Statistical Methods*, (23rd Revised ed.). New Delhi: Vikas Publishing House PVT Ltd.

MODULE 3

UNIT 1 PROBABILITY THEORY I

CONTENTS

1.0 INTRODUCTION

In our daily activities or events we come across reasons to look for ways of predicting events (or happenings) before they actually occur. For instance, we may want to ask a number of potential voters whether they favour a certain candidate or not. The foundation of probability is usually ascribed to the 17th-century French Mathematicians *Blaise Pascal* and *Pierre de Fermat*, but Mathematicians such as *Gerolamo Cardano* had made important contributions to its development even before this time. Mathematical probability began in an attempt to answer certain questions arising in games of chance, such as how many times a pair of dice must be thrown before the chance that a six will appear is 50-50.

Usually, many unpleasant events occur in life simply because we could not predict its occurrence as to guide against it. A man once beaten by rain said, *"If I had known I would have taken an umbrella with me in the morning, now it is raining."* It is because of this fact of man wanting to measure these levels of daily uncertainty that this branch of Mathematics known as *Probability Theory* or *Theory of Probability* was introduced.

2.0 OBJECTIVES

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

- explain some basic terms as they relate to probability
- solve problems of experimental probability
- solve problems of theoretical probability.

3.0 MAIN CONTENT

3.1 Definition of probability

The probability of an event, E, happening or occurring is the measure of the possibility that the event E will occur in any one trial or experiment carried under a defined condition. The probability of an event E occurring is usually denoted by Pr(E).

3.2 Types of Probability

Probability can broadly be classified into two types namely:

a) Experimental (or empirical) probability
b) Theoretical (or classical) probability

3.2.1 Experimental (or Empirical) Probability

When several experiments are actually carried out under the same condition through which an empirical result or outcome is obtained then the probability of such an event is what is called *Experimental Probability*.

Example

Take a die of six faces and throw it 50 times. Put your outcome in the tabular form as shown below:

Die	1	2	3	4	5	6
No of occurrence	10	8	7	5	8	12

After the experiment is completed, one can then begin to calculate the probability of obtaining may be a number (1) or a number (5) or a number (6).

3.2.2 Theoretical (or Classical) Probability

When the probability of an event can be obtained from an already existing result, from what is likely to occur or from some already known information, then we call such a theoretical or classical probability.

3.3 Definition of some Basic Terms in Probability

1. Experiment

In Mathematics, the word *experiment* is used to describe a process that will generate numeric data, which can be used for analysis. Essentially, we can classify experiment into two groups namely:

a. Deterministic Experiment

In this type of experiment, the results of the experiment are not subject to chance if repeated severally under the same condition. For example, measuring the length of a particular straight line would give the same result always.

b. Random Experiment

In this type of experiment, the results are always subject to chances and may change if the experiment is repeated even under the same condition. For example, gambling; casinos; rolling a die; shuffling cards; tossing a coin; and voting. In probability theory we are mainly concerned with *random experiments* since it is the only one from which measures of uncertainty is inferred.

2. An Outcome

An outcome is the result of any experiment carried out under a well-defined condition in probability. For instance, if in a throw of a coin a head shows up then the head (H) is an outcome.

3. Equally Likely Outcomes

We say that the outcomes of an experiment are *equally likely* if each of the outcomes has an equal chance of occurring. For instance, the head and tail of a fair die are equally likely outcomes of the experiment of tossing a coin since they both have equal chances of occurring.

Politically, the results of each aspirant for a post are equally likely outcomes since they all have equal chances of been elected.

3.4 Sample Space and Sample Points

The *sample space* of a random experiment is the collection or set of all possible distinct outcomes of the experiment. In a set theory, a sample space is called *universal set*. For example:

i, Throwing a fair coin once would have a sample space of {H,T}.
ii. Throwing two fair coins once would have a sample space of {HH, HT, TH, TT}.
iii. In a primary election of party-AB, the sample space of the party's candidates *(assuming there are four of them)* is {Bayo, Yusuf, Eze, Tzong}.

A *sample point* on the other hand is a single point (or element) in the sample space. In the example above, each candidate Bayo, Yusuf, Eze and Tzong is a sample point.

We shall denote a sample space by *S* so that the number of sample points (or elements) in the sample space will be devoted by n(*S*)

3.5 An Event Space

An *event* is simply a subset of the sample space. Essentially, an event is a collection of sample points characterised by some descriptive properties or features. The total number of events that can be gotten from a sample space containing *n* number of sample points is always 2^n. So a sample space *S* = {H, T} would have the subsets (events): {H}, {T}, {H,T} and { }. *(Where { } is the empty set)*

We shall however denote an event by *E* and the number of outcomes (or element) in *E* by n(*E*).

3.6 Axioms of Probability

1. If we denote the number of outcomes in an event space *E* by n(*E*) and the number of outcomes in a sample space *S* by n(*S*), then the probability of an event *E* denoted by Pr(*E*) is defined as: Pr(*E*) $= \frac{n(E)}{n(S)}$
2. Let Pr(a) be the probability of an event that is impossible to occur then

$$\Pr(a) = 0.$$

3. Let Pr(b) be the probability of an event that is very certain to occur, then

   Pr (b) = 1.

4. Thus, the probability of any event E is therefore a number that satisfies the inequalities: $0 \leq \Pr(E) \leq 1$. That is at no occasion must the probability of an event be less than 0 or greater than 1.

5. If the probability of an event *E* occurring is denoted by Pr(*E*) and the probability of an event *E* not occurring is denoted by Pr(E'), then

$$\Pr(E) + \Pr(E') = 1.$$

Solved Examples

1. Two fair coins are tossed together, find the probability that at least a tail shows up.

Solution

S = {HH, HT, TH, TT}, n(S) = 4
E = {HT, TH, TT}, n(E) = 3

Hence, Pr(at least a tail shows up) = $\frac{n(E)}{n(S)} = \frac{3}{4}$

2. If an unbiased die is thrown, find the probability that the number which show up is less than or equal to 4.

Solution

S = {1, 2, 3, 4, 5, 6}, n(S) = 6
E = {1, 2, 3, 4}, n (E) = 4
Hence Pr (a number less than or equal to 4) = $\frac{n(E)}{n(S)} = \frac{4}{6} = \frac{2}{3}$

3. A tin box contains 9 *black*, 6 *blu*e and 10 *red* marbles. If a marble is picked from the box at random what is the probability that it is a red marble?

Solution

Given 9 black, 6 blue and l0 red marbles.
n(S) = 9 + 6 + 10 = 25
n(E) = 10
Hence, Pr(a no. ≤ 4) = $\frac{10}{25} = \frac{2}{5}$

4. If two fair dice are rolled together what is the probability of obtaining (a) a total of 4, (b) a number greater than 4 (c) a number divisible by 4 (d) a number which is a multiple of 5.

Solution

Sample space for 1st die = {1, 2, 3, 4, 5, 6}
Sample space for 2nd die = {1, 2, 3, 4, 5, 6}

The two dice would be

	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

Total →

	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

(a) n(S) = 36, n(E) = n(total of 4) = 3. Hence, Pr(total of 4) = $\frac{3}{36} = \frac{1}{12}$

(b) n(S) = 36, n(E) = n(total no. < 4) = 3. Pr(total no. < 4) = $\frac{3}{36} = \frac{1}{12}$.

Hence pr (total nos. >4) = 1 - $\frac{1}{12} = \frac{11}{12}$

(c) n(S) = 36, n(E) = n(total no of divisible 4) = 9.

Hence, Pr(total of no. divisible by 4) = $\frac{9}{36} = \frac{1}{4}$

(d) n(S) = 36, n(E) = n(total of no. that are multiples of 5) = 7.

Hence, Pr(total of nos. which are multiples of 5) = $\frac{7}{36}$

5.

	Ward-1	Ward -2	Ward-3
Male	23	44	27
Female	16	21	33

The table above show the number of party members according to their wards and sexes. If a party member is selected at random from the table, find the probability that the person is (a) a male (b) a female (c) a member from ward-2 (d) a male and from ward-3 (e) a female from ward-1 (f) a member from ward-7.

Solution

n(S) = 23 + 44 + 27 + 16 + 21 + 33 = 164
n(male) = 23 + 44 + 27 = 94,   n(female) = 16 + 21 + 33 = 70

(a) Pr(a male) $= \frac{n(male)}{n(S)} = \frac{94}{164} = \frac{47}{82}$

(b) Pr (a female) = $\frac{n(female)}{n(S)} = \frac{70}{164} = \frac{35}{82}$

(c) Pr(a member from ward-2) = $\frac{65}{164}$

(d) Pr(a male and ward-3) = $\frac{n(male\ from\ ward-3)}{n(S)} = \frac{27}{164}$

(e) Pr(a female and ward -1) = $\frac{n(female\ and\ ward-1)}{n(S)} = \frac{16}{164} = \frac{4}{41}$

(f) Pr(a member from ward-7) = $\frac{n(ward-7)}{n(S)} = \frac{0}{164}$ = 0.

6. The result of a primary election of party AA of five aspirants in ascending order is: 22, 26, 32, 55, 71 and the result of a primary election of party AB of five aspirants in ascending order is: 12, 19, 28, 34, 88. If two aspirants one from AA and one from AB are picked, what is the probability that

(a) the sum of their votes is greater than 80
(b) the sum of their votes is less than 60
(c) the vote of aspirant from party AA is greater than the vote of the aspirant from party AB
(d) the vote of aspirant from party AB is greater than the vote of the aspirant from AB.

Solution

Sample space for party AA = {22, 26, 32, 55, 71}
Sample space for party AB = {12, 19, 28, 34, 88}

For the two parties would be:

	22	**26**	**32**	**55**	**71**
12	34	38	44	67	83
19	41	55	51	74	90
28	50	54	60	83	99
34	56	60	66	89	105
88	110	114	120	143	159

Total

	22	**26**	**32**	**55**	**71**
12	34	38	44	67	83
19	41	45	51	74	90
28	50	54	60	83	99
34	56	60	66	89	105
88	110	114	120	143	159

(a) n(S) = 25, n(E) = n(sum of votes > 80) = 11.
Hence, Pr(sum of votes > 80) = $\frac{11}{25}$

(b) n(S) = 25, n(E) = n(sum of votes < 60) = 9.
Hence, Pr(sum of vote ,60) = $\frac{9}{25}$

(c) n(S) = 25, n(E) = n(vote of AA > vote of AB) = 15
Hence, Pr(vote of AA > vote of AB) = $\frac{15}{25} = \frac{3}{5}$

(d) n(S) = 25, n(E) = n(vote of AB > vote AA) = 10

Hence Pr(vote of AB > vote of AA) = $\frac{10}{25} = \frac{2}{5}$

SELF ASSESSMENT EXERCISE

1. If the probability of chief Adio winning an election is $\frac{1}{3}$, what is the probability that he loses the election? .
2. In a single toss of three fair dice, what is the probability of:
   a) The three dice showing the same number.
   b) The three dice showing different number.
   c) The three dice showing prime number.

4.0 CONCLUSION

In this unit, you have learnt the different types of probability and the types of experiments that will generate numerical values. Another important thing you have learnt is the *basic axioms of probability*, which are very useful in finding solutions to problems related to predicting uncertainty.

5.0 SUMMARY

In this unit, you have learnt the following:

1 Probability of an event E occurring is the measure of the possibility that the event E will occur in any one trial or an experiment carried out under a defined condition.
2 Types of probability are *experimental* and *theoretical* probability.
3 Types of experiment are *deterministic* and *random* experiments.
4 The probability of an event E is defined as Pr(E).
5 The probability of an event that would never occur is 0.
6 The probability of an event that is certain to occur is 1.
7 The probability of an event satisfies the inequalities: $0 \leq \Pr(E) \leq 1$.
8 If E denotes an event occurring and E' denotes an event not occurring, then, $\Pr(E) + \Pr(E') = 1$, so $\Pr(E) = 1 - \Pr(E')$ or $\Pr(E') = 1 - \Pr(E)$. E' is called complementary of event E.

6.0 TUTOR–MARKED ASSIGNMENT

1. In a single toss of two fair dice, what is the probability of the difference in scores being 2?

2.

	Science	Agriculture	Soc. science	Arts
Bolaji	66	108	98	102
Adisa	58	82	112	88
Abdul	74	18	64	54

The above table shows the result of an election from four faculties for the seat of the SUG president of a reputable university in Nigeria. The three aspirants are: Bolaji, Adisa, and Abdul.

(a) If the sample space of the distribution above is represented by *S*, *f*ind n(*S*).
(b) If one of the candidates is picked at random, what is the probability that it is

i. Adisa ii. Abdul
iii. His name does not start with A

(c) Find the probability that Bolaji does not win the election
(d) Find the probability that Bolaji wins in faculty of Agriculture
(e) Find the probability that Abdul wins in faculty of Arts

3. Distinguish with relevant examples between the following pairs of terms:

a. Deterministic and random experiments.
b. Experimental and theoretical probabilities.
c. Sample point and sample space.

7.0 REFERENCES/FURTHER READING

Adekola, A.O. (1993). *Further Mathematics*. Macmillan Nigeria Publishers Ltd, Nigeria.

Gupta, C. B. & Vijay Gupta (2004). *An Introduction to Statistical Methods.*

Goodman, A. & Talbert, J.F. (1999). *Additional Mathematics for West Africa.* Longman Group Ltd.

Stroud, K. A. (1995). *Engineering Mathematics* (4th ed.). Lagos, Nigeria: Macmillan Press Ltd.

UNIT 2 PROBABILITY THEORY II

CONTENTS

1.0 INTRODUCTION

In the first unit, we learnt the definitions of some terms that would help to understand the concept of probability. Afterwards we learnt the different ways of solving probability problems of singular events. In this unit, we shall be dealing with two or more events that happen (or occur) together or simultaneously. Such events are called combined events.

2.0 OBJECTIVES

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

- explain the concept of combined events
- use the addition law of probability to solve relevant problems
- use the multiplication law of probability to solve relevant problems.

3.0 MAIN CONTENT

3.1 Combination of Events

Sometimes we carry out two or more random experiments that eventually lead to two or more events occurring. When this happens, it is possible that the first event affects or influences the second event or not. To settle with the problem of one event affecting the other, there are two important laws that can be applied depending *strictly* on the type of events in question.

3.2 Mutually Exclusive and Mutually Non-Exclusive Events

a) Mutually Exclusive Events

Two or more events are *mutually exclusive* if they cannot occur (or happen) at the same time. For example,

i. In the event of tossing a coin, the H (head) and T (tail) can never occur at the same time.
ii. In the event of throwing a dice, a (1) and a (5) can never occur at the same time.
iii. Constitutionally, the event of electing both Mr. A and Mr. B as the president of Nigeria in the same election can never happen since only one president can be elected at a time. In statistics, this event is mutually exclusive.

In *set theory*, this is an example of a disjoint set or a distinct set.

b) Mutually Non-Exclusive Events

Two or more events are said to be mutually non-exclusive events if it is possible for them to occur (or happen) at the same time. For example:

i. If a dice is thrown, the event of getting an odd number could be getting a 1, 3 or 5 which are all odd numbers.
ii. The event of getting a presidential aspirant from a particular state could be more than one.

3.3 The Addition Law of Probability

If two events E_1 and E_2 are mutually exclusive, then in set notation,
$n(E_1 \text{ OR } E_2) = n(E_1) + n(E_2) + 0$ (since $n(E_1 \cap E_1) = (0)$.

So that the probability of (E_1 OR E_2) is defined by $Pr(E_1 \cup E_2) = P(E_1) + P(E_2)$.

If however events E_1 and E_2 are mutually non-exclusive events, then
$P_r(E_1 \cup E_2) = P(E_1) + P(E_2) - P_r(E_1 \cap E_1)$.

Note: (i) The union "$\cup$" property takes the "OR" phrase which is usually reflected in problems where *addition law* can be used.

(ii) The law can be extended to situations of more than two events.

Example

1. A six-sided fair dice is rolled once. Find the probability that the side that shows up is
   (a) either a 1 or a 5
   (b) either an even number or a prime number
   (c) either an odd number or a perfect square.
2. In an election, the probability that Mr P, Mr Q and Mr R win the election is: $\frac{1}{5}, \frac{1}{4}$ and $\frac{1}{3}$ respectively, calculate the probability that;
   (a) Mr Q did not win
   (b) Mr P or Mr R wins.

Solution

1. $S = \{1, 2, 3, 4, 5, 6\}$, $n(S) = 6$
   (a) $E_1 = \{1\},\ n(E_1) = 1, E_2 = \{5\}, n(E_1) = 1$
   $Pr(E_1$ OR $E_2) = Pr(E_1) + Pr(E_2)$ (mutually exclusive events)
   So $Pr(E_1$ OR $E_2) = Pr(1$ OR $5) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$
   (b) $E_1 = \{2,4,6\},\ n(E_1) = 3, E_2 = \{2,3,5\},\ n(E_2) = 3,$
   So, $(E_1 \cap\ E_2) = \{2\},\ n(E_1 \cap E_2) = 1$
   And
   $Pr(E_1\ OR\ E_2) = Pr(E_1) + P_r(E_2) - Pr\ (E_1\ AND\ E_2)$
   (mutually non-exclusive events)
   So $Pr\ (E_1\ OR\ E_2) = Pr(Even\ OR\ Prime)\ = \frac{3}{6} + \frac{3}{6} - \frac{1}{6} = \frac{5}{6}$
   (c) $E_1 = \{1,3,5\},\ n(E_1) = 3,\ \ E_2 = \{1,4\},\ n(E_2) = 2$
   $Pr(E_1\ OR\ E_2) = Pr(E_1) + Pr(E_2)$ (Mutually exclusive events)
   So $Pr\ (E_1\ OR\ E^2) = Pr(odd\ OR\ perfect\ square)\ \frac{3}{6} + \frac{2}{6} = \frac{5}{6}$
2. Given that $Pr(Mr\ \ P\ wining)\ = \frac{1}{5}$
   $Pr(Mr\ Q\ wining)\ = \frac{1}{4}$
   $Pr(Mr\ R\ wining)\ = \frac{1}{3}$ then
   (a) If $Pr(Mr\ Q\ wining)\ = \frac{1}{4}$ then
   $Pr(Mr\ Q\ not\ wining)\ = 1 - \frac{1}{4} = \frac{8}{15}$
   (b) *Pr(Mr P wining OR Mr R wining) = Pr(Mr P)+ Pr(Mr R)* (Mutually exclusive events)
   So, *Pr*(Mr P wins OR Mr R wins) = $\frac{1}{5} + \frac{1}{3} = \frac{8}{15}$

3.4 Independent and Dependent Events

a) **Independent Events:** Two or more events are said to be *independent* if the occurrence of one event does not affect or influence the occurrence of the second event. For example, assuming there are 4 blue balls, 5 black balls and 6 red balls in a bag. If we pick a ball the first time, then replace the ball and again pick a ball the second time, then the type of ball we pick the first time does not, in anyway have any effect or influence on the second ball that was picked. Hence we say that the two outcomes are independent.

b) **Dependent Events:** Two or more events are said to be *dependent* if the occurrence of one event does affect or influence the occurrence of the second event. For example, assuming there are 4 blue balls, 5 black balls, and 6 Red balls in a bag. If we pick a ball the first time, then without replacing it we again pick a ball the second time, then the type of ball we picked the first time does, in all ways, have effect or influence on the second ball that was picked. Hence we say that the two outcomes are *dependent*.

3.5 The Product (or multiplication) law of Probability

If two events E_1 and E_2 are independent then $Pr(E_1 \cap E_2) = P_r(E_1) \times P_r(E_2)$

Note: (i) The intersection "$\cap$" property takes the "AND" phrase which is usually reflected in problems where the product law of probability can be used.

(ii) The law can be extended to situation of more than 2 events.

Example

1. A box contains 4 black and 6 white identical beads. If a girl picks two beads randomly one after the other *with replacement*. What is the probability that

i. The first bead is black and the second bead is white
ii. Both beads are of the same colour
iii. The first bead is white and the second is black
iv. Both beads are of different colours

Solution

Given that blue = 4 and white = 6.

(i) $Pr(black\ then\ white) = Pr(black\ AND\ white)$

$= Pr(black) \times Pr(white) = \frac{4}{10} \times \frac{6}{10} = \frac{2}{5} \times \frac{3}{5} = \frac{6}{25}$

(ii) $Pr(both\ same\ colour)$

$= Pr(black\ AND\ black\ )\ OR\ P(white\ AND\ white)$

$= Pr(black\ AND\ black\ ) + \ Pr(white\ AND\ white)$

$= Pr(black\ ) \times Pr(black\ ) + \ Pr(white)$

$\times Pr(white)$

$= \left(\frac{4}{10} \times \frac{4}{10}\right) + \left(\frac{6}{10} \times \frac{6}{10}\right) = \left(\frac{2}{5} \times \frac{2}{5}\right) + \left(\frac{3}{5} \times \frac{3}{5}\right)$

$= \frac{4}{25} + \frac{9}{25} = \frac{13}{25}$

(iii) $Pr(white\ then\ black\ ) = Pr(white\ AND\ black)$

$= Pr(white) \times Pr(black) = \frac{6}{10} \times \frac{4}{10} = \frac{3}{5} \times \frac{2}{5} = \frac{6}{25}$

(iv) $Pr(different\ colour)$

$= Pr(black\ AND\ white\ )\ OR\ Pr(white\ AND\ black)$

$= Pr(black) \times Pr(white\ ) + Pr(white) \times Pr(black)$

$= \left(\frac{4}{10} \times \frac{6}{10}\right) + \left(\frac{6}{10} \times \frac{4}{10}\right) = \left(\frac{2}{5} \times \frac{3}{5}\right) + \left(\frac{3}{5} \times \frac{2}{5}\right)$

$= \frac{6}{25} + \frac{6}{25} = \frac{12}{25}$

2. A box contains 4 black and 6 white identical beads. If a girl picks two beads randomly one after the other *without replacement*. What is the probability that:

i. The first bead is black and the second bead is white?
ii. The first bead is white and the second bead is black
iii. Both beads are of the same colour?
iv. Both beads are of different colours?

Solution

Given that black = 4 and white = 6

(i) $Pr(black\ then\ white\ ) = \ Pr(black\ AND\ white)$
$= Pr(black) \times Pr(white)$
$$= \frac{4}{10} \times \frac{6}{9} = \frac{2}{5} \times \frac{2}{3}$$
$$= \frac{4}{15} (since\ 1st\ bead\ is\ not\ replaced)$$

(ii) $Pr(both\ same\ colour)$
$= Pr(black\ AND\ black\ )\ OR\ P(white\ AND\ white)$
$= Pr(black\ AND\ black\ ) + \ Pr(white\ AND\ white)$

$$= Pr(black\ ) \times Pr(black\ ) + \ Pr(white) \times Pr(white)$$
$$= \left(\frac{4}{10} \times \frac{3}{9}\right) + \left(\frac{6}{10} \times \frac{5}{9}\right) = \left(\frac{2}{5} \times \frac{1}{3}\right) + \left(\frac{3}{5} \times \frac{5}{9}\right)$$
$$= \frac{2}{15} + \frac{1}{3} = \frac{7}{15}$$
$(since\ 1st\ bead\ is\ replaced)$

(iii) $Pr(\ white\ then\ black\ ) = \ Pr(black\ AND\ white)$
$= Pr(white) \times Pr(black)$
$$= \frac{6}{10} \times \frac{4}{9} = \frac{3}{5} \times \frac{4}{9}$$
$$= \frac{4}{15} (since\ 1st\ bead\ is\ not\ replaced)$$

(iv) *Pr(both of different colours) = Pr(black AND white) OR Pr(white AND black)*
$= Pr(\ black) \times Pr(\ white) + \ Pr(white) \times Pr(black)$
$$= Pr\left(\frac{4}{10} \times \frac{6}{9}\right) + \left(\frac{6}{10} \times \frac{4}{9}\right)$$
$$= \frac{8}{15}\ (since\ 1st\ bead\ is\ replaced)$$

3. The probability that Jack, James and John are elected in their different constituencies are respectively $\frac{1}{3}, \frac{2}{5}\ and\ \frac{3}{8}$. Find the Probability that:

i. The three of them are elected
ii. None of them is elected
iii. Only one of them is elected

Solution

Given that $Pr(Jack) = \frac{1}{3}, Pr(James) = \frac{2}{5}$ $and\ Pr(John) = \frac{3}{8}$

i. *Pr(the three of them elected) = Pr(Jack AND James AND John)*

$$= Pr(Jack) \times Pr(James) \times Pr(John) = \frac{1}{3} \times \frac{2}{5} \times \frac{3}{8} = \frac{1}{20}$$

ii. *Pr(none of them is elected) = Pr(NOT Jack AND NOT James AND NOT John)*

$$= Pr(not\ Jack) \times Pr(not\ James) \times Pr(not\ John)$$

$$if\ Pr(Jack) = \frac{1}{3}, then\ Pr(not\ Jack) = 1 - \frac{1}{3} = \frac{2}{3}$$

$$if\ Pr(James) = \frac{2}{5}, then\ Pr(not\ James) = 1 - \frac{2}{5} = \frac{3}{5}$$

$$if\ Pr(John) = \frac{3}{8}, then\ Pr(not\ John) = 1 - \frac{3}{8} = \frac{5}{8}$$

$$so, Pr(none\ of\ them\ is\ elected) = \frac{2}{3} \times \frac{3}{5} \times \frac{5}{8} = \frac{1}{4}$$

iii. $Pr(only\ one\ is\ elected) = Pr(\ Jack,\ not\ James, not\ John)$

$OR\ Pr(not\ Jack, James, not\ John)\ OR\ Pr(not\ Jack, not\ James, John)$

$$= Pr(Jack) \times Pr(not\ James) \times Pr(not\ John) + Pr(not\ Jack) \times Pr(James) \times Pr(not\ John) + Pr(not\ Jack) \times Pr(not\ James) \times Pr(Jack)$$

$$= \left(\frac{1}{3} \times \frac{3}{5} \times \frac{5}{8}\right) + \left(\frac{2}{3} \times \frac{2}{5} \times \frac{5}{8}\right) + \left(\frac{2}{3} \times \frac{3}{5} \times \frac{3}{8}\right)$$

$$= \frac{1}{8} + \frac{1}{6} + \frac{3}{20} = \frac{41}{120}$$

SELF ASSESSMENT EXERCISE

In a re-run election in three different states in Nigeria, the probability that Mr. A, Mr. B and Mr. C win in their various states are given as: $\frac{1}{5}, \frac{2}{3} and\ \frac{2}{5}\ respectively.$ what is the probability that:

i. all of then lose in their respective states
ii. at least one of them wins
iii. at least two of them win
iv. at most two of them win

4.0 CONCLUSION

In this unit, you have learnt the methods of combining two or more events based on the manner of their selection. We have also learnt how to apply the *addition law* and *product law* to solving problems of mutually exclusive events and independent events.

5.0 SUMMARY

From this unit, we learnt that:

1. Two events E_1 and E_2 are said to be *mutually exclusive* if the two events cannot occur at the same time.
2. Two events E_1 and E_2 are said to be *independent* if the occurrence of E_1 does not affect or influence the occurrence of E_2.
3. If events E_1 and E_2 are mutually exclusive events then $Pr(E_1 \cup E_2) = Pr(E_1 OR\ E_2) = Pr(E_1) + Pr(E_2)$, this is called *addition law* of probability.
4. If events E_1 and E_2 *mutually nonexclusive* events then
$Pr(E_1 \cup E_2) = Pr(E_1 OR\ E_2) =$
$Pr(E_1)+P_r(E_2) - Pr(E_1\ AND\ E_2)$
This is called *addition law* of probability.
5. If events E_1 and E_2 are *independent* events then
$Pr(E_1 \cap E_2) = Pr(E_1 AND\ E_2) = Pr(E_1) \times Pr(E_2)$.
This is called the product law of probability.

6.0 TUTOR-MARKED ASSIGNMENT

1. A folder contains 7 blue biros, 6 red biros and 3 black biros. If a man selected three biros without replacement, find the probability that:

   i. The three biros are of the same colour
   ii. The three biros are of different colours
   iii. At least two are of the same colour.
   iv. At most two are of the same colour

2. The table below shows the result of a gubernatorial election of four political parties PA, PB, PC and PD.

Political party	**PA**	**PB**	**PC**	**PD**

Result	2,035	1,224	5,094	1,700

(a) If a party is selected at random, what is the probability that it is PD?

(b) If a party is randomly selected, what is the probability that its result is more than 1,700?

(c) If two parties are selected at random, without replacement, find the probability that:

(i) the parties is either PA or PD

(ii) The parties are both PA and PD.

3. In Kankaro village, every eligible voter belongs to political parties AA, AB, AC or AD. The population of each political party is given in the table below.

	AA	**AB**	**AC**	**AD**
Male	56	48	94	66
Female	32	12	18	22

If two voters are selected at random from the village, find the probability that:

i. They are both males or both females
ii. They are not of similar sex.
iii. They are from the same party
iv. They are from different parties
v. They are both males and from party AB.
vi. They are of different sexes from different parties.

7.0 REFERENCES/FURTHER READING

Adekola, A. O. (1993). *Further Mathematics*. Nigeria: Macmillan Nigeria Publishers Ltd.

Goodman, A. & Talbert, J.F. (1999). *Additional Mathematics for West Africa.* Longman Group LTD.

Tuttuh - Adegun, M.R; Sivasubramaniam, S. and Adegoke, R. (2004). *Further Mathematics Project.* Ibadan, Nigeria: NPS Educational Publishers Limited.

Stroud, K.A. (1995). *Engineering Mathematics*, (4th ed.). Macmillan Press Ltd.

UNIT 3 PERMUTATION THEOREM

CONTENTS

1.0 INTRODUCTION

Sometimes we may ask ourselves "how many ways can I arrange 5 items in a row?" The problem of knowing the number of ways a number of items can be arranged gave rise to the concept of *permutation*. Permutation simply means *arrangement*. So we can always know beforehand the number of ways a number of items could be arranged.

2.0 OBJECTIVES

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

- calculate the factorial of a number
- explain the concept of permutation
- calculate the number of arrangement of dissimilar objects
- calculate the number of arrangement of similar objects
- calculate the number of arrangement of circular objects
- calculate the number of arrangement of objects given a condition.

3.0 MAIN CONTENT

3.1 Definition and Concept of Factorial of a Number

The *factorial* of a number n ($n \geq 0$) usually denoted by $n!$ is defined as
$n! = 1 \times 2 \times 3 \times 4 \times \ldots \times (n-2) \times (n-1) \times n$

So that;
$$0! = 1 = 1$$
$$1! = 1 = 1$$
$$2! = 1 \times 2 = 2$$
$$3! = 1 \times 2 \times 3 = 6$$
$$4! = 1 \times 2 \times 3 \times 4 = 24$$
$$5! = 1 \times 2 \times 3 \times 4 \times 5 = 120$$

Example 1

Evaluate the following

(1) $8!$ (2) $2! \times 5!$ (3) $\frac{(7-2)!}{3!}$

(4) $2(8-3)!$ (5) $\frac{6!(7-3)!}{3!4!2!}$

Solution

(1) $8! = 1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 = 40{,}320$

(2) $2! \times 5! = (1 \times 2) \times (1 \times 2 \times 3 \times 4 \times 5) = 2 \times 120 = 240$

(3) $\frac{(7-2)!}{3!} = \frac{5!}{3!} = \frac{5 \times 4 \times 3!}{3!} = 5 \times 4 = 20$

(4) $2(8-3)! = 2(5!) = 2 \times 120 = 240$

(5) $\frac{6!(7-3)!}{3!4!2!} = \frac{6!4!}{3!4!2!} = \frac{6!}{3!2!} = \frac{6 \times 5 \times 4 \times 3!}{3!2!} = \frac{120}{2} = 60$

3.2 Permutation as Arrangement of Objects

The *permutation* (or arrangement) of *n* distinct objects is its factorial. So 5 distinct objects can be arranged in 5! =120 ways, and 10 distinct objects can be arranged in 10! = 3,628,800 ways. We may want to permute (arrange) *r* distinct objects out of *n* distinct objects such that consideration is given to the *order of arrangement*. The number of permutation of *n* distinct objects taken *r* distinct objects at a time is termed

n permutation *r* usually denoted r; ${}^{n}P_{r}$ or ${}_{n}P_{r}$ or $P_{n,r}$

${}^{n}P_{r}$ is defined as $\frac{n!}{(n-r)!}$

Example 2

Evaluate the following;

(1) $^{10}P_4$ (2) $^{n}P_0$ (3) $^{8}P_2 \times {}^{5}P_2$ (4) $^{4}P_3 \div {}^{4}P_2$
(5) $^{5}P_2 - {}^{6}P_3$

Solution

(1) $^{10}P_4 = \frac{10!}{(10-4)!} = \frac{10!}{6!} = \frac{10\times9\times8\times7\times6!}{6!} = 10 \times 9 \times 8 \times 7 = 5{,}040$

(2) $^{n}P_0 = \frac{n!}{(n-0)!} = \frac{n!}{n!} = 1$

(3) $^{8}P_2 \times {}^{5}P_2$
$= \frac{8!}{(8-2)!} \times \frac{5!}{(5-2)!} = \frac{8!}{6!} \times \frac{5!}{3!} = \frac{8\times7\times6!}{6!} \times \frac{5\times4\times3!}{3!} = 56 \times 20 = 1120$

(4) $^{4}P_3 \div {}^{4}P_2$
$= \frac{4!}{(4-3)!} \div \frac{4!}{(4-2)!} = \frac{4!}{1!} \div \frac{4!}{2!} = 4 \div \frac{4\times3\times2!}{2!} = 4! \div 4 \times 3 = 2$

(5) $^{5}P_2$ - $^{6}P_3$
$= \frac{5!}{(5-2)!} - \frac{6!}{(6-3)!} = \frac{5!}{3!} - \frac{6!}{3!} = \frac{5\times4\times3!}{3!} - \frac{6\times5\times4!3}{3!} = 20 - 120 =$
-100

Example 3

(1) Find the number of permutations of the letters of the word JAMES. (*Ans*. There are 5 letters; hence no. of permutation is 5! = 120 ways)

(2) Ten voters cards ate labelled A, B, C, D, E, F, G, H, I, J. Taking any four cards at a time, find the number of ways the cards can be arranged. (*Ans.* Number of permutation = $^{10}P_4$ $= \frac{10!}{(10-4)!} = \frac{10!}{6!} = 5{,}040 ways$)

(3) In a country there are 20 political parties. In the next election any 5 political parties' name are to be printed on the 4 segments of its ballot pages. How many ways can the arrangement be done? (*Ans:* no. of arrangement = $^{20}P_5$ =

$$\frac{20!}{(20-5)!} = \frac{20!}{15!} = \frac{20 \times 19 \times 18 \times 17 \times 16 \times 15!}{15!}$$
$$= 20 \times 19 \times 18 \times 17 \times 16 = 1{,}860{,}480$$

3.3 Permutation of Identity Objects

Consider the permutation of the letters of word MATHEMATICS. Note that there are;

i. 2 identical letters M
ii. 2 identical letters A
iii. 2 identical letters T

So that the identical letters M can be arranged in 2 times = 2!
And the identical letters A can be arranged in 2 times = 2!
And the identical letters T can be arranged in 2 times = 2!
So that the eleven (11) letters of the word MATHEMATICS can be arranged in

$\frac{11!}{2!2!2!}$ ways.

Example 4

Find the number of ways to permute the letters of the word

(1) ELIMINATION (2) NOUN (3) FANTASTIC

Solution

1. No of ways of permutation ELIMINATION = $\frac{11!}{3!2!}$ (for 3'I's is and 2'N's)
2. No of ways of permutation NOUN = $\frac{4!}{2!}$ (for 2'N's) = 12 ways.
3. No of ways of permutation FANTASTIC = $\frac{9!}{2!2!}$ = $\frac{9!}{4}$ =90,720 ways.

3.4 Cyclic Permutation

The concept of cyclic permutation is about arrangement of objects about a circle. Consider finding the number of permutations of *n* persons round a circular table. Since a round table has neither a beginning nor an end, we will fix one object and permute the remaining (*n*-1) objects about the fixed one object.

However if the circle can be turned over, then the number of permutation is $\frac{(n-1)!}{2}$ ways.

Example 5

1. In how many ways can the executive committee members of 6 be seated round a table? (*Ans*: no. of ways = (6-1)! = 5! = 120 ways)
2. A bunch of keys consisting of 10 distinct keys are fixed into a round ring. In how many ways can these keys be arranged in this ring.

   (*Ans*: no. of ways = $\frac{(10-1)!}{2}$ = $\frac{9!}{2}$ = 181440 (since the bunch of key can be turned over)

3.5 Conditional Permutation

When modifications or restrictions are placed on the order of arrangement of the distinct objects then the permutation is called a *conditional permutation*. In such a case we must put the condition, restriction or modification into consideration when calculating the no. of arrangement.

Example 6

(1) In how many ways can the letters of the word COMMITTEE be arranged if all doubled letters must always be together? (*Ans*: if doubled letters must be together then COMMITTEE has C | O | MM | I | TT | EE - i.e. 6 objects to be arranged, so no. of permutation is 6! = 720 ways.

(2) Calculate the no. of ways the digits 1, 2, 3 and 4 can be arranged to give a number greater than 3000. (*Ans*: for the number to be greater than 3000 then the numbers must start with digit 3 or 4 hence we fix digit 3 and permute the remaining 3 digits. So, no. of ways = (1×3!) + (1×3!) = 6 + 6 = 12.)

4.0 CONCLUSION

In this unit, you have learnt how to calculate the factorial of a number. Also you have learnt how to calculate the number of ways a given number of objects can be arranged.

5.0 SUMMARY

From this unit we learnt that;

1. $n! = n \times (n-1) \times (n-2) \times \ldots \times 3 \times 2 \times 1$
2. $0! = 1; 1! = 1$

3. ${}^{n}P_r = \frac{n!}{(n-r)!}$ ${}^{n}P_0 = 1$ and ${}^{n}P_1 = n$
4. The permutation of n distinct objects taking r objects at a time is defined as ${}^{n}P_r$.
5. The permutation of n objects such that r_1, r_2, and r_3, are each identical, then number of ways of permuting the n-objects is $\frac{n!}{r_1!r_2!r_3!}$
6. The permutation of n objects round a circle that cannot be turned over is $(n-1)!$ If however it can be turned over then it is $\frac{(n-1)!}{2}$.
7. We should always put all conditions or restrictions attached to a problem into consideration when calculating the permutation of objects.

6.0 TUTOR-MARKED ASSIGNMENT

1. In how many ways can each of the letters of the following words be arranged?
   (a) COMMITTEE (b) NIGERIA
   (c) TELECOMMUNICATION (d) OSHOGBO
   (e) EXCELLENCE (f) ACCOMMODATION

2. Evaluate the following;

   (a) $\frac{6!(7-3)!}{3!4!2!}$ (b) $\frac{8!}{(8-4)!4!}$ (c) $\frac{2n!\times(2n+1)}{(2n+1)!}$

3. In a quiz, in how many ways can the 1st, 2nd and 3rd prizes be awarded to 25 contestants?

7.0 REFERENCES/FURTHER READING

Adekola, A.O. (1993). *Further Mathematics*. Nigeria: Macmillan Nigeria Publishers Ltd.

Goodman, A. & Talbert, J.F. (1999). *Additional Mathematics for West Africa*. Longman Group Ltd.

Tuttuh - Adegun, M.R; Sivasubramaniam, S. and Adegoke, R. (2004). *Further Mathematics Project.* Ibadan, Nigeria: NPS Educational Publishers Limited.

Stroud, K.A. (1995). *Engineering Mathematics*. (4th ed.). Macmillan Press Ltd.

UNIT 4 COMBINATION THEOREM

CONTENTS

1.0 INTRODUCTION

Suppose we have 5 cards labelled A, B, C, D, and E, then we can arrange the cards in threes i.e. ${}^5P_3$ = 60 ways. Since the order of arrangement is relevant and important, then ABC, BAC and CAB are two different arrangements. Now if the order of arrangement is irrelevant and less important then all arrangements such as ABC BAC, CAB, ACB, etc are all the same and shall be counted as one type of arrangement. This type of arrangement in which the order of arrangement is less important is called *Combination*.

2.0 OBJECTIVES

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

- evaluate combinational expressions
- recognise a problem of selection or combination.

3.0 MAIN CONTENT

3.1 Definition of Combination

While permutation is about the arrangement of objects taking r ($r \leq n$) objects at a time (with interest in order of arrangement), combination is about selecting a certain number of r objects from a set of n objects, in this case, the order of arrangement is of no interest nor importance, thus if there are n numbers of objects, then the number of ways of selecting r objects is defined as:

$$ {}^nC_r = \frac{n!}{(n-r)!r!} $$

${}^{n}C_r$ is also denoted by: ${}^{n}C_r$ or $\binom{n}{r}$.

3.2 Evaluation of *n* Combination *r*

Example 1

Evaluate the following;

(a) ${}^{20}C_{15}$ (b) ${}^{20}C_5 \div {}^{20}P_5$ (c) If ${}^{n}P_3 = 4({}^{n}C_5)$ find the value of n.

Solution

1. ${}^{20}C_{15} = \frac{20!}{(20-15)!15!} = \frac{20!}{5!15!} = \frac{20\times19\times18\times17\times16\times15!}{5!15!} = \frac{15504}{5}$

2. ${}^{20}C_5 \div {}^{20}P_5 \ \frac{20!}{(20-5)!5!} \div \frac{20!}{(20-5)!} = \frac{20!}{15!5!} \times \frac{15!}{20!} = \frac{1}{5!}$

3. ${}^{n}P_3 = 4\ {}^{n}C_5$

$\frac{n!}{(n-3)!} = \frac{4n!}{(n-5)!5!}$

$\frac{1}{(n-3)!} = \frac{4}{(n-5)!5!}$

$\frac{1}{(n-3)(n-4)(n-5)!} = \frac{4}{(n-5)!\,5!}$

$\frac{1}{(n-3)(n-4)} = \frac{4}{5!}$

$4(n - 3)(n - 4) = 5! = 120$

$(n - 3)(n - 4) = 30$

$n^2 - 7n + 12 - 30 = 0$

$n^2 - 7n - 18 = 0$

$(n - 9)(n + 2) = 0$

$n - 9 = 0$ or $n + 2 = 0$

$n = 9$ or $n = -2$

But n cannot be negative so $n = 9$ is the solution.

3.3 Combination or Selection

Example 2

1. In how many ways can 6 questions be chosen from 10 optional questions?

Solution

No. of selections is ${}^{10}C_6 = \frac{10!}{(10-6)!6!} = \frac{10\times9\times8\times7\times6!}{4!\times6!} = 210$ ways

2. A committee of 3 men and 2 women is to be chosen from 7 men and 5 women. In how many ways can the selection be made?

Solution

No. of selection is $^7C_3$ AND $^5C_2$ = $^7C_3 \times {}^5C_2 = 35 \times 10 = 350$ways

3. A box contains 12 male voter's cards and another box contains 10 female voter's cards. Find the number of ways of selecting 5 male cards and 5 female cards if 2 particular female cards must be selected.

Solution

Selecting male cards = $^{12}C_5 = 792$ways
Selecting the remaining 3 female cards = $^8C_3 = 56$ways
Selection = $792 \times 56 = 44352$ways

4. A committee of 6 men and 4 women is to be formed from 10 men and 8 women. (a) In how many ways can this committee be formed? (b) If a particular woman will not serve on the committee with a particular man, find the number of ways this committee can be formed.

Solution

(a) No. of committee = $^{10}C_6 \times {}^8C_4 = 210 \times 70 = 14700$
(b) No. of committee with a particular woman not serving with a particular man = (no. of possible committee) - (no. of committee involving the two of them) = 14700 - ($^9C_5 \times {}^7C_3$)
= 14700 - 4410 = 10290ways

5. Find the number of ways of choosing four party agents from five men and three women if at least one woman must be in the selection.

Solution

At least one woman = (1W AND 3M) OR (2W AND 2M) OR (3W AND 1M)
$= ({}^3C_1 \times {}^5C_3) + ({}^3C_2 \times {}^5C_2) + ({}^3C_3 \times {}^5C_1)$
= 30 + 30 + 5 = 65ways

6. An opposition group of five is to be formed from five men and four women. Calculate the number of ways this can be done if the number of men must be greater than the number of women.

Solution

For number of men greater than number of women

= (5M AND 0W) OR (4M AND 1W) OR (3M AND 2W)
= $({}^5C_5 \times {}^4C_0) + ({}^5C_4 \times {}^4C_1) + ({}^5C_3 \times {}^4C_2)$
= 1 + 20 + 60 = 81 ways.

SELF ASSESSMENT EXERCISE

1. Five delegates out of five men and five women are to represent an electoral body at a national debate. Find the number of ways the delegates can be selected if (a) it must consist of 2 men and 3 women (b) the delegates must have at least one person of each sex.
2. In how many ways can 5 secretaries, 4 clerks, 3 drivers and 2 cleaners be chosen from 10 secretaries, 9 clerks, 8 drivers and 7 cleaners?

4.0 CONCLUSION

In this unit, you have learnt how to calculate the number of ways a given number of objects can be arranged such that order of arrangement is irrelevant.

5.0 SUMMARY

From this unit we learnt that;

$n! = n \times (n-1) \times (n-2) \times ... \times 3 \times 2 \times 1$

$0! = 1;\ 1! = 1$

Combination means selection or choosing *r* distinct objects from *n* distinct objects.

${}^nC_r = \frac{n!}{r!(n-r)!} \quad (r \le n)$, , ${}^nC_0 = 1$ and ${}^nC_1 = n$

$${}^nC_r = \frac{{}^nP_r}{r!} = \frac{\frac{n!}{(n-r)!}}{r!} = \frac{n!}{(n-r)!} \div r! = \frac{n!}{(n-r)!} \times \frac{1}{r!} = \frac{n!}{(n-r)!r!}$$

6.0 TUTOR-MARKED ASSIGNMENT

1. The INEC Chairperson is to select 2 regional coordinators for each of the 6 political zones in Nigeria. If there are 15 qualified people for these post, calculate the number of ways he can do the selection, (assuming any person can be taken to any region).
2. In a country there were 30 political parties. Later a concession was reached to form two mega parties of at least 13 political parties in each mega party. In how many ways can this be done assuming all the 30 parties must belong to any of the two mega parties?
3. A caretaker committee of five is to be selected from six men and four women to run the affair of a local government. How many different committees can be formed if (a) any person can be taken (b) the number of men on the committee must be greater than the number of women?

7.0 REFERENCES/FURTHER READING

Adekola, A.O. (1993). *Further Mathematics.* Nigeria: Macmillan Nigeria Publishers Ltd.

Goodman, A. & Talbert, J.F. (1999). *Additional Mathematics for West Africa*. Longman Group Ltd.

Tuttuh - Adegun, M.R; Sivasubramaniam, S. and Adegoke, R. (2004). *Further Mathematics Project.* Ibadan, Nigeria: NPS Educational Publishers Limited.

Stroud, K.A. (1995). *Engineering Mathematics*. (4th ed.). Macmillan Press Ltd.

UNIT 5 PROBABILITY DISTRIBUTIONS

CONTENTS

1.0 INTRODUCTION

Over the years, some certain experiments were found to have just two possible types of outcome. A voter's card can only be valid or invalid. A member can only belong to a party or not. In addition, tossing a coin as many times as possible would only show a *head* or a *tail* (two possible outcomes). In an examination, a student can only *pass* or *fail*. All the given examples above have a common attribute of yes/no, true/false, good/bad, pleasant/unpleasant or success and failure. Owing to this special attribute, statisticians over the years had developed some mathematical models (functions). These models can be used to calculate the probability of the illustrated natural occurrences above. The models are called *Probability distribution functions*.

2.0 OBJECTIVES

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

- distinguish between discrete and continuous data
- use Binomial distribution formula to solve probability problems
- use Poisson distribution formula to solve probability problems
- identify which of the two distribution functions is to be used to solve a given problem.

3.0 MAIN CONTENT

The two probability distributions that we shall discuss in this course are:

(a) Binomial distribution
(b) Poisson distribution

We want to say at this junction that both the Binomial distribution and the Poisson distribution are discrete distributions. A discrete distribution contains discrete data i.e. whole numbers or integers e.g. 0, - 49, 2, 1000, -11, etc. while continuous data are decimal numbers e.g. 0.54, 24.4, 6.042, etc.

3.1 The Binomial Distribution

Suppose in an experiment, there are just two possible outcomes i.e. *success* and *failure*. Let the probability of a success be *p* and let the probability of a failure be *q*. Obviously, from our knowledge of probability theory, $p + q = 1$, $p = 1 - q$ and $q = 1 - p$. Now if the experiment is carried out in *n* number of times under the same condition and there are *x* number of successes then there should be (*n* - *x*) number of failures. So that the probability of a success is defined by the binomial distribution:

$Pr(x) = C_x p^x q^{n-x}$, where ^nC_x means *n* combination *x*.

Examples

1. A fair coin is thrown 5 times. Find the probability of a tail showing up 3 times.

Solution

Here probability of success i.e. getting a tail is $\frac{1}{2}$

Probability of failure, i.e. getting not a tail is $q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2}$

Number of trials is $n = 5$
Number of success is $x = 3$
And $Pr(x) = {}^nC_r\, p^x q^{n-x}$

$Pr(x = 3) = {}^5C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{5-3} = \frac{5}{16}$

2. In eight throws of a fair dice, what is the probability of obtaining a six?

Solution

$p = \frac{1}{6}, \quad q = 1 - p = 1 - \frac{1}{6} = \frac{5}{6}, \quad n = 8, \quad x = 6$

And $Pr(x) = {}^nC_r\, p^x q^{n-x}$

So $Pr(x = 6) = {}^8C_6 \left(\frac{1}{6}\right)^6 \left(\frac{5}{6}\right)^{8-6} = \frac{25 \times 7 \times 4}{6^8} = \frac{700}{6^8} = 0.00042$

3. In a political group of 12 members the probability of picking a member from any one of the six geo-political zones in Nigeria is $\frac{2}{3}$. If some members are to represent the group in a National meeting, find the probability that:

(a) 8 members are chosen
(b) At most 8 members are chosen
(c) At least 8 members are chosen

Solution

(a) $p = \frac{2}{3},\ q = 1 - p = 1 - \frac{2}{3} = \frac{1}{3}\quad ,\quad n = 12,\ x = 8$

So, $Pr(x = 8) = {}^{12}C_8$

$\left(\frac{2}{3}\right)^{12}\left(\frac{1}{3}\right)^{12-8} = 2^{12} \times 11 \times 5 \times 9 \times \left(\frac{1}{3}\right)^{16}$

$= 0.0471$

(b) If at most 8 members are chosen then

$p = \frac{2}{3},\qquad q = \frac{1}{3},\qquad n = 12,$

$x \leq 8\ (i.e.\ x = \{0,1,2,3,4,5,6,7,8\})$

$Pr(x \leq 8) = Pr(x = 0) + Pr(x = 1) + Pr(x = 2) + Pr(x = 3) + \cdots + Pr(x = 8)$

This seems cumbersome so we have to apply a little common sense.

Note that $Pr(x \leq 8) = 1 - Pr(x > 8)$.

i.e. $Pr(x = 8) = 1 - Pr(x = 9) + Pr(x = 10) + Pr(x = 11) + Pr(x = 12)$

Now, $\Pr(x = 9) = C_9^{12}\left(\frac{2}{3}\right)^{12}\left(\frac{1}{3}\right)^{12-9} = 0.0628$

$\Pr(x = 10) = C_{10}^{12}\left(\frac{2}{3}\right)^{12}\left(\frac{1}{3}\right)^{12-10} = 0.0565$

$\Pr(x = 11) = C_{11}^{12}\left(\frac{2}{3}\right)^{12}\left(\frac{1}{3}\right)^{12-11} = 0.0308$

$\Pr(x = 12) = C_{12}^{12}\left(\frac{2}{3}\right)^{12}\left(\frac{1}{3}\right)^{12-12} = 0.0077$

So, $\Pr(x \leq 8) = 1 - \{0.0628 + 0.0565 + 0.0308 + 0.0077\}$

$= 1 - 0.1578 = 0.8422$

(c) If at least 8 members are chosen then

$$p = \frac{2}{3}, q = 1 - \frac{2}{3} = \frac{1}{3}, \quad n = 12,$$
$$x \geq 8 \ (i.e. x = \{8,9,10,11,12\}).$$
$$Pr(x \geq 8) = Pr(x = 8) + Pr(x = 9) + Pr(x = 10)$$
$$+ Pr(x = 11) + Pr(x = 12)$$

From (a) and (b) above

$$Pr(x \geq 8) = 0.0471 + 0.0628 + 0.0565 + 0.0308$$
$$+ 0.0077 = 0.2049$$

4. In the production of ballot papers for a gubernatorial election, it was found that there are 3 defective printings out of every 15 papers. If five papers are picked at random, find the probability that:

(a) 2 are defective
(b) At most 2 are defective
(c) At least 2 are defective.

Solution

(a) If 2 are defective,

$$p = \frac{3}{15} = \frac{1}{5}, \ q = 1 - p = \frac{4}{5}, \ n = 5, \ x = 2.$$
$$Pr(x = 2) = C_2\left(\tfrac{1}{5}\right)^{5}{}_{5}\left(\tfrac{4}{5}\right)^{5-2} = 0.0016$$

(b) If at most 2 are defective

$$p = \frac{3}{15} = \frac{1}{5}, \ q = 1 - p = \frac{4}{5}, \ n = 5, \ x \leq 2(x = \{0,1,2\}).$$
$$Pr(x \leq 2) = Pr(x = 0) + Pr(x = 1) + Pr(x = 2).$$
$$Pr(x = 0) = C_0\left(\tfrac{1}{5}\right)^{5}{}_{5}\left(\tfrac{4}{5}\right)^{5-0} = 0.0001$$
$$Pr(x = 1) = C_1\left(\tfrac{1}{5}\right)^{5}{}_{5}\left(\tfrac{4}{5}\right)^{5-1} = 0.0007$$
$$Pr(x = 2) = C_2\left(\frac{1}{5}\right)^{5}{}_{5}\left(\frac{4}{5}\right)^{5-2} = 0.0016$$

So $Pr(x \leq 2) = (0.0001) + (0.0007) + (0.0016) = 0.0024$

(c) If at least 2 are defective

$$P = \frac{1}{5}, \qquad q = \frac{4}{5}, \qquad n = 5, \qquad x \geq 2(i.e. x = \{2,3,4,5\}$$

Note that $Pr(x \geq 2) = 1 - Pr(x < 2)$

So $Pr(x \geq 2) = 1 - \{Pr(x = 0) + Pr(x = 1)\}$

$$= 1 - \{0.0001 + 0.0007\} = 0.9976$$

3.2 The Poisson Distribution

Recall that both the binomial distribution and the Poisson distribution are discrete distributions. Now what makes the difference between the two?

Over the years statisticians found out that when the number of trials (i.e. *n*) is very large (say $n \geq 30$) and the probability of a success (i.e. *p*) is relatively very small then the binomial distribution may not be the best distribution to be used. A better option is the Poisson distribution.

The Poisson distribution is defined as:

$$Pr(x) = \frac{\lambda^x e^{-\lambda}}{x!} (x \geq 0)$$

$where\ \lambda = np$ (*n* is the number of trials and *p* is the probability of success and x is the number of success).

$e = 2.7183\ (5\ s.f)$

Examples

1. At a political party convention, an issue was put up to be deliberated upon. In taking decision it was found that one out of a thousand members supported the decision. If there are 5,000 members at the convention find the probability that:

(a) 5 members supported the decision
(b) At most 2 members supported the decision
(c) At least 3 members supported the decision.

Solution

(a) For 5 members to support the decision

$$p = 1\ out\ of\ a\ thousand = \frac{1}{1000},\ x = 5\ , n = 5000,$$

$$\lambda = np = 5000 \times \frac{1}{1000} = 5$$

(Note that *p* is very small and *n* is very large)

And $Pr(x) = \frac{\lambda^x e^{-\lambda}}{x!}$

So $Pr(x = 5) = \frac{5^5 e^{-5}}{5!} = \frac{5^5 \times 2.7183^{-5}}{5!} = 0.1755$

(b) For at most 2 members to support the decision

$$p = \frac{1}{1000}, \quad n = 5000, \quad \lambda = np = 5,$$
$$x \leq 2 (i.e.x = \{0,1,2\})$$

So $Pr(x \leq 2) = Pr(x = 0) + Pr(x = 1) + Pr(x = 2)$.

$$= \left(\frac{5^0 \times 2.7183^{-5}}{0!}\right) + \left(\frac{5 \times 2.7183^{-5}}{1!}\right) + \left(\frac{5^2 \times 2.7183^{-5}}{2!}\right)$$

$Pr(x \leq 2)$ = 0.0067 + 0.0337 + 0.0842 = 0.1246

(c) For at least 3 members to support the decision

$p = \frac{1}{1000}, n = 5000, \lambda = np = 5, x \geq 3$ $(i.e.x = \{3,4,5, \dots \dots ,5000\})$

Note that $Pr(x \geq 3) = 1 - Pr(x \leq 2)$ = 1 - 0.1246

2. There are 15 wards in a local government. In each ward, there are 10 political aspirants. If the probability of any aspirant winning a primary election is $\frac{1}{10}$. Calculate the probability that five of the aspirants win their election.

Solution

$n = 15 \times 10 = 150$ Aspirants, $p = \frac{1}{10}$, $x = 5$, $\lambda = np = 150 \times \frac{1}{10} = 15$

And, $Pr(x) = \frac{\lambda^x e^{-\lambda}}{x!} (x \geq 0)$.

$Pr(x = 5) = \frac{15^5 \times 2.7183^{-15}}{5!} = 0.0019$

3. In selecting 100 election officials that will conduct elections in some wards across four states, 10% were found to be illiterates. What is the probability that

(a) 2 persons selected are illiterates.
(b) No person selected is an illiterate.
(c) Not more that 2 officials selected are illiterates.

(d) Not less than 2 officials selected are illiterates.

Solution

(a) $n = 100, p = 10\% = \frac{10}{100} = \frac{1}{10}, \lambda = np = 100 \times \frac{1}{10} = 10, x = 2$

And $Pr(x) = \frac{\lambda^x e^{-\lambda}}{x!}$

So, $Pr(x = 2) = \frac{10^2 \times 2.7183^{-10}}{2!} = 0.0023$

(b) For no illiterate is selected

$n = 100, \quad p = \frac{1}{10}, \quad \lambda = 10, \quad x = 0$

$Pr(x = 0) = \frac{10^0 \times 2.7183^{-10}}{0!} = 0.00005$

(c) For not more than 2 illiterates

$n = 100, \ p = \frac{1}{10}, \ \lambda = 10, \ x \ngtr 2 \ (i.e. x \leq 2, x = \{0,1,2\})$

$Pr(x \ngtr 2) = Pr(x = 0) + Pr(x = 1) + Pr(x = 2)$

$= \left(\frac{10^0 \times 2.7183^{-10}}{0!}\right) + \left(\frac{10^1 \times 2.7183^{-10}}{1!}\right) + \left(\frac{10^2 \times 2.7183^{-10}}{2!}\right)$

$= 0.00005 + 0.00045 + 0.00227 = 0.0028$

(d) For not less than 2 illiterates,

$n = 100, \ p = \frac{1}{10}, \ \lambda = 10, \ x \nless 2 \ (i.e. x \geq 2, x = \{2,3,4, \ldots\})$

Note that $Pr(x \nless 2) = 1 - \{\Pr(x = 0) + \Pr(x = 1)\}$.

$= 1 - \{0.00005 + 0.00045\} = 0.9995$

4. If 99.5% of the population of a village belongs to a political party XYZ. Find the probability that 2 out of 5000 villagers *does not* belong to the party XYZ.

Solution

$n = 5000, \quad q = 99.5\%,$

$p = 100\% - 99.5\% = 0.5\% = \frac{5}{10}\% = \frac{1}{200}$

$\lambda = np = 5000 \times \frac{1}{200} = 10$

$\Pr(x = 2) = \frac{10^2 \times 2.7183^{-10}}{2!} = 0.0023$

5. After supplying 10,000 ballot papers, 1 out of 500 ballot papers were found to be defective. What is the probability that any 5 randomly selected ballot papers are defective?

Solution

$$n = 10{,}000 \quad p = \frac{1}{500}, \qquad \lambda = np = 10{,}000 \times \frac{1}{500} = 20, x = 5$$

$$\Pr(x = 5) = \frac{20^5 \times 2.7183^{-20}}{5!} = 0.000055$$

SELF ASSESSMENT EXERCISE

1. As part of electoral materials, a box containing 10 black ink pads, 8 blue ink pads and 7 red ink pads got its pads shuffled in transit. If 5 pads are picked at random, find the probability that:

(a) The 5 are red ink pads
(b) The 5 are black ink pads
(c) The 5 are neither red nor black ink pads.

2. If 5% of words in a memorandum are misspelled, what is the probability that 50 words selected at random will have not less than 3 words misspelled?

4.0 CONCLUSION

In this unit, you have learnt that there are two types of data namely *discrete data* and *continuous data*. Also you have learnt that with given number of success of an event and the probability of an event occurring, we can decide which of the distributions is suitable in calculating the probability of an *x* number of trials of an accent.

5.0 SUMMARY

In this unit you have learnt the following:

1. How to use the two probability distribution functions:

(i) Binomial distribution and
(ii) Poisson distribution

2. That the binomial distribution and the Poisson distribution are discrete distributions.
3. The binomial distribution is defined as:

$$\Pr(x) = C_x p^x q^{n-x}$$

4. The Poisson distribution is defined as:

$$Pr(x) = \frac{\lambda^x e^{-\lambda}}{x!}(x \geq 0).$$

5. The parameters *n* is the number of independent trials in an experiment

p is the probability of success
q is the probability of failure
${}^nC_x$ means *n* combination *x* i.e. $\frac{n!}{x!(n-x)!}$

$$\lambda = np$$

x = number of successes in *n* trials

6.0 TUTOR-MARKED ASSIGNMENT

1. 15 aspirants emerged from a ward. If the probability that an aspirant wins an election is (0.8). Find the probability that any four aspirants from the ward win the election.
2. To select an electoral official, he must pass an aptitude test that contains 10 questions each for two sections A and B. The questions contain Yes or No answers. For a candidate who did the examination by guessing, what is the probability that he got 10 questions in all correctly?
3. 100 INEC officials were recommended for a course overseas. Before they can be allowed, they were to be injected with a particular vaccine. If 1 out of 20 officials are found to be resistant to the vaccine, what is the probability that (5) officials selected at random would be found to be resistant to the vaccine?

7.0 REFERENCES/FURTHER READING

Stroud, K. A. (1995) *Engineering Mathematics* (4th ed.). Macmillan Press Ltd.

Tuttuh - Adegun, M. R; Sivasubramaniam, S. and Adegoke, R. (2003). *Further Mathematics Project 2*. NPS Educational.

MODULE 4

UNIT 1 NATURE AND IMPORTANCE OF STATISTICAL INQUIRIES

CONTENTS

1.0 INTRODUCTION

Statistical inquiries have to do with the use of investigation to determine the facts of a case using statistical method of analysis. A bus contains 18 seats. The owner sells tickets in advance for seats in the bus, and he realises that 12% of the people who buy tickets do not show up each day. In order for him to maximise profits, the owner decides to sell 20 tickets each day. The owner wishes to find out how many times per week the bus will be overbooked. The word “**inquiry**” by definition means to investigate, to search for knowledge, or to find out information about something. Statistical Inquiry is a search that relates to data and numerical facts which can be measured quantitatively. For accuracy in measurement, it is important that before any research or inquiry is conducted, the purpose and scope of such an inquiry should be clearly defined.

2.0 OBJECTIVES

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

- discuss the meaning of Statistical inquiry
- discuss the importance of statistical inquiry
- define the terms; Sampling, Population, Random Sampling etc
- distinguish between the various types of sampling
- enumerate and explain the laws of sampling procedure.

3.0 MAIN CONTENT

3.1 Some Basic Concepts

3.1.1 Population

Population is the entire group of individuals or objects under consideration. According to Allan G. Blueman, population means all subjects (human or otherwise) under study. Since some populations can be very large, researchers cannot use every single subject.

3.1.2 A Sample

This is a portion, piece, or segment that represents the whole. For instance, product samples are provided by product manufacturers in the hope that they encourage future sales. We also take samples of body fluids to help determine the condition of the body. Chemical substances or minerals in an area are taken as samples to determine the composition of the area surrounding a simple site. In the same vein, demographic and statistical samples help in understanding and quantifying mathematical probabilities, trends, or relationships within a group. We may also take samples of sounds or pictures as measurements at a given moment, which are expressed digitally, and can be used in various communications systems.

3.1.3 Sample Survey

A sample survey is a method of collecting data. It involves obtaining answers to a number of questions from different individuals, usually in the form of a survey. The reliability of a survey's results depends on whether the sample of people from which the information has been collected is free from bias and sufficiently large enough.

3.1.4 Random Sampling

In order for a sample to be a random sample, every member of the population must have an equal chance of being selected. When a sample is chosen at random from a population, it is said to be an unbiased sample. In other words, the sample, for the most part, is representative of the population. On the other hand, if a sample is selected incorrectly, it may be a biased sample when some kind of systematic error has been made in the selection of the data.

3.2 Importance of Statistical Inquiries and Sampling

A sample is used to get information about a population for several reasons. In many cases, sampling remains the only way to determine something about the population. Some of these reasons are discussed below:

- It saves time of the researcher. For example, a candidate for a presidential seat may wish to determine his chances in an election. A sample poll could be carried out. Using the regular staff and field interviews of a professional statistician would take only one or two days. This of course will not be possible if dealing with the whole voting population. Sampling reduces the cost of research.

- Sampling enables the researcher to get information with ease. For example, if a person's blood is to be tested for Human Immuno-deficiency Virus (HIV), the virus that causes AIDS, the researcher cannot analyse every single drop of blood without killing the person. Also, if the truthfulness of claims over a product is to be investigated, the researcher does not have to subject each and every one of the products one after the other to analysis to draw a conclusion but rather it is sufficient to randomly select a sample and carry out necessary test(s) to investigate the claims of the manufacturer.

- It helps the researcher to get more detailed information about the subject in question. When a sample of a population is surveyed, the researcher can conduct in depth interviews by spending more time with each person, thus getting more information about the subject. This is not to conclude that the smaller the sample, the better the result. In general, large samples give more reliable information if correct sampling techniques are used.

3.3 Some Incorrect Methods of Sampling

In random sampling, the basic requirement is that for a sample of size 'n', all possible samples of this size must have an equal chance of being selected from the population. But some researchers employ some incorrect methods of sampling, this will be discussed below:

- Asking questions from people on the street is an incorrect method commonly used by news reporters. This method does not meet the basic requirement for simple random sampling, since not all possible samples of the population have an equal chance of being selected. Majority of the people are likely to be at home or at work when the interview is being conducted and therefore do not have a chance of being selected.
- Another incorrect way of asking questions for research purposes is through the media (radio or television), asking the listeners and viewers to call the station to give their responses or opinions on a matter. This method is not random, since only those who watch or listen to the program could be part of the respondents. In addition, only those who feel strongly for or against the issue may respond. Another wrong method of sampling, closely related to this, is when people are asked to respond through mail. Only those who are concerned and who have the time are likely to respond.

Not all the methods discussed above meet the requirements of random sampling, since not all possible samples of a specific population have an equal chance.

3.4 Types of Sampling

3.4.1 Simple Random Sampling

One of the most widely used types of sampling is the simple random sampling method, where each item or person in the population has the same chance of being included. For example, in order to use Simple Random Sampling for a population of 900 employees of a plastic industry, a sample of 60 employees is selected from that population. One way of ensuring that every employee in the population has a chance of being chosen is to first write the name of each one on a small slip of paper and deposit all of the slips in a container. After they have been mixed thoroughly, the selection is made by drawing out of the box without looking at it. The process is repeated 60 times. The major limitation of random sampling is that it is not quite suitable where the population under survey is large. If the population is extremely large it is time consuming to number and select the sample elements.

3.4.2 Cluster and Focus Group Sampling

Cluster and focus group sampling involve selecting some known members of a population for study e.g. Footballers, Prostitutes, Petroleum dealers, adolescents, women, senior citizen, etc. This is a biased method of sampling, because there is a deliberate focus on some groups to the exclusion of others. When a particular stratum is selected for study, then, either cluster or focus group sampling is said to have been used in selecting the sample points. Data collection for content analysis – content counting – seems to have fallen into this category. It involves selecting a news or communication medium and counting the number of times, a given event (e.g. corruption activity) is reported. This technique is especially useful in the analysis of newspaper publications, for example, reports on, accidents, thefts and the like.

3.4.3 Random Systematic Sampling

If a population having N members is arranged in ascending or descending order from 1 to N, to select a sample of (n) units from it using the methods of Random Systematic Sampling, we take the first sample by a random procedure and thereafter, the next K^{th} item is selected (where K is the regularity or interval at which we select sample). For example, if the first randomly selected sample is number 3 and K is 10^{th} item after number 3, then, the items or samples we shall come up with are 3, 13, 23, 33, etc., all selected systematically.

3.4.4 Stratified Random Sampling

A sampling procedure that first stratifies the population using any of the methods of Cluster and Focus Group (FG) sampling procedures described above and thereafter selects samples from each stratum using random sampling techniques is called *Stratified Random Sampling.* For instance, income can be used to classify population into high, medium and low-income groups and proportional samples are selected by random procedures from each stratum based on their relative population.

3.4.5 Hybrid Sampling Methods

A number of sampling techniques combine random and non-random procedures. These include; stratified random and random systematic sampling techniques, among others.

SELF ASSESSMENT EXERCISE

1. List the advantages and disadvantages of random sampling
2. List and explain the incorrect methods that are often used to obtain a sample
3. In your opinion, which of the sampling method(s) studied provides the best sample to represent a population?

4.0 CONCLUSION

The study of statistics is highly valuable and instructive because it sheds light upon the social relations of society, and imparts information regarding its political condition. In this chapter, we outlined the major methods and procedures for selecting a representative sample from a given population. Such techniques include stratified random sampling and (random) systematic sampling techniques. The social condition of a country is of more vital importance than its political condition, as the maintenance of peace and good order depends more upon the former, than on the latter. Hence, the importance of statistical inquiries cannot be over emphasised.

5.0 SUMMARY

In this unit, we have been able to examine the nature and importance of Statistical Inquiries. We gave some basic definitions, which will help us to understand the concept in question thoroughly. The subject of sampling was also examined in detail. We outlined the major methods and procedures for selecting a representative sample from a given population as well. It is shown that the most reliable way of selecting a representative sample is the use of random sampling. Other non-random sampling methods discussed include, cluster and focus group sampling, content counting and systematic sampling. It was shown that some hybrid sampling techniques combine elements of random and non-random sampling techniques. Such techniques include stratified random sampling and (random) systematic sampling techniques.

6.0 TUTOR-MARKED ASSIGNMENT

1. Define the following terms
   a. Population
   b. Sample Survey
2. Why are samples used in statistics?
3. Discuss the importance of statistical inquiries.

7.0 REFERENCES/FURTHER READING

Nachmias, D. and Nachmias, C. (1976). *Research Methods in the Social Sciences.* London.

Okoro, E. (2002). *Quantitative Techniques in Urban Analysis.* Ibadan: Kraft Books Ltd.

Show, G. and Wheeler, D. (1985). *Statistical Techniques in Geographical Analysis.* New York: John Wiley and Sons.

Gupta, C.B. and Vijay, Gupta (1973). *An Introduction to Statistics Methods*, (23rd revised Edition). USB Publishers Distribution PVT Ltd.

William Menden Hall; Robeit J. Beaver, and Barbara M. Beaver (2003). *Probability and Statistics.* United States of America.

UNIT 2 BASIC RESEARCH METHODOLOGY (I)

CONTENTS

1.0 INTRODUCTION

The selection of a primary method of investigation of a given problem is a key consideration for the investigator. There must be a thorough knowledge of the basic methods of research to guide the researcher in the choice of a primary research method to be used in solving his problem. All factual knowledge, which is ascertained by research, may be classified in terms of three areas of time:

(1) The past: What has been? Historical Research Method;
(2) The present: What is now occurring? Normative – Survey Research Method, experimental research Method;
(3) The future: What probably will be? Prognostic or predictive Research Method. *See E.C Osuala*

2.0 OBJECTIVES

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

- define the concept of search
- discuss the different types of research methodology
- enumerate the advantages and disadvantages of research methodology.

3.0 MAIN CONTENT

3.1 Understanding Basic Concepts

3.1.1 Research

Research can be defined to be search for knowledge or any systematic investigation to establish facts. The primary purpose for applied research is to discover, interpret, and develop methods and systems for the advancement of human knowledge on a wide variety of scientific matters of our world and the universe. The term research can also be defined as a series of activities carried out systematically to find answers to identified problems in a particular discipline. It is the application of the scientific approach to the study of a problem. It is also used to describe an entire collection of information about a particular subject.

3.1.2 Scientific Research

It relies on the application of the scientific method, a harnessing of curiosity. This type of research provides scientific information and theories for the explanation of the nature and the properties of the world around us. It makes applications that are as practical as possible.

Scientific research is funded by public authorities, by charitable organisations and by private groups, including many companies. Scientific research can be subdivided into different classifications according to their academic and application disciplines.

In scientific studies, accurate analysis of data achieved by the use of standardized statistical methods is critical to determining the validity of empirical research. Statistical formulas such as regression, uncertainty coefficient, t-test, chi-square, and various types of ANOVA (analyses of variance) are fundamental to forming logical, valid conclusions. If empirical data reaches significance under the appropriate statistical formula, the research hypothesis is supported. If not, the null hypothesis is supported, meaning no effect of the independent variable(s) was observed on the dependent variable(s).

It is important to understand that the outcome of empirical research using statistical hypothesis testing is never foolproof. It can only support a hypothesis, reject it, or do neither. These methods yield only probabilities. Among scientific researchers, empirical evidence (as distinct from empirical research) refers to objective evidence that appears the same regardless of the observer. For example, a thermometer will not display different temperatures for each individual who observes it. Temperature, as measured by an accurate, well-calibrated thermometer, is empirical evidence. By contrast, non-empirical evidence is subjective, depending on the observer. Following the previous example, observer A might truthfully report that a room is warm, while observer B might truthfully report that the same room is cool, though both observe the same reading on the thermometer. The use of empirical evidence negates this effect of personal (i.e., subjective) experience. Ideally, empirical research yields empirical evidence, which can then be analysed for statistical significance or reported in its raw form. The following information is intended to give the reader some general guidance about planning a basic research effort in their organisation. The rest of the information in the section presents an overview of methods used in business: how to apply them; how to analyse, interpret, and report results.

3.1.3 Research Processes

Generally, research is understood to follow a certain structural process. Though step order may vary depending on the subject matter and researcher, the following steps are usually part of most formal research, both basic and applied:

- Formation of the topic
- Hypothesis

- Conceptual definitions
- Operational definition
- Gathering of data
- Analysis of data
- Test, revising of hypothesis
- Conclusion, iteration if necessary

3.1.4 Hypothesis

A hypothesis is used to make predictions that can be tested by observing the outcome of an experiment. If the outcome is inconsistent with the hypothesis, then the hypothesis is rejected. However, if the outcome is consistent with the hypothesis, the experiment is said to support the hypothesis. This careful language is used because researchers recognise that alternative hypotheses may also be consistent with the observations. In this sense, a hypothesis can never be proven, but rather only supported by surviving rounds of scientific testing and, eventually, becoming widely thought of as true (or better, predictive), but this is not the same as it having been proven. A useful hypothesis allows prediction and within the accuracy of observation of the time, the prediction will be verified. As the accuracy of observation improves with time, the hypothesis may no longer provide an accurate prediction. In this case, a new hypothesis will arise to challenge the old, and to the extent that the new hypothesis makes more accurate predictions than the old, the new hypothesis will supplant the old.

3.1.5 Research methods

The goal of the research process is to produce new knowledge, which takes three main forms (although, as previously discussed, the boundaries between them may not be clear-cut):

- Exploratory research, which structures and identifies new problems
- Constructive research, which develops solutions to a problem
- Empirical research, which tests the feasibility of a solution using empirical evidence

3.1.6 Research Funding

Most funding for scientific research comes from two major sources, corporate organisations (through research and development departments) and government (primarily through universities and in some cases through military contractors). Many senior researchers (such as group leaders) spend a considerable amount of their time applying for grants for research funds. These grants not only provide resources to carry out the research, they also serve as a source of merit.

3.2 Considerations to Design Your Research Approach

1. For what purposes is the research being done, i.e., what do you want to be able to decide as a result of the research?
2. Who are the audience for the information from the research, e.g., fund-providers/ bankers, upper management, employees, customers, etc?
3. What kinds of information are needed to make decisions or to enlighten your intended audience? For instance, do you need information to really understand a process, the customers who buy certain products; strengths and weaknesses of the product or service or programme; benefits to customers; how the product or service or programme failed some customers and why, etc.?
4. From what sources should the information be collected, e.g., employees, customers, groups of employees or customers, certain documentation, etc.?
5. How can that information be collected in a reasonable manner, e.g., questionnaires, interviews, examining documentation, observing staff and/or clients in the programme, conducting focus groups among staff and/or clients, etc.
6. When the information is needed what is the period within which it must be collected?
7. What resources are available to collect the information? Secondary research can come from either internal or external sources.

3.3 Overall Goal in Selecting Research Methods

The overall goal in selecting basic business research method(s) is to get the most useful information to key decision makers in the most cost-effective and realistic fashion.

1. What information is needed to make current decisions about a product or programme?
2. How much of this information can be collected and analysed in a cost-effective and practical manner, e.g., using questionnaires, surveys and checklists?
3. How accurate will the information be (reference the above table for disadvantages of methods)?
4. Will the methods get all the needed information?
5. What additional methods should and could be used if additional information is needed?

6. Will the information appear as credible to decision makers, e.g., to bankers, fund-providers or top management?
7. Will the nature of the audience conform to the methods, e.g., will they fill out questionnaires carefully, engage in interviews or focus groups, let you examine their documentations, etc.?
8. Who can administer the methods now or is training required?
9. How can information be analysed?

Note that, ideally, the researcher uses a combination of methods, for example, a questionnaire to quickly collect a great deal of information from a lot of people, and then interviews to get more in-depth information from certain respondents to the questionnaires. Perhaps case studies could then be used for more in-depth analysis of unique and notable cases, e.g., those who benefited or not from the programme, those who quit the programme, etc.

3.4 Four Levels of Research Results

There are four levels of information that can be gathered from customers or clients, including the following:

1. Reactions and feelings (feelings are often poor indicators that your service made lasting impact)
2. Learning (enhanced attitudes, perceptions or knowledge)
3. Changes in skills (applied the learning to enhance behaviours)
4. Effectiveness (improved performance because of enhanced behaviours)

Usually, the farther your research results get down the list, the more useful is your research results. Unfortunately, it is quite difficult to reliably get information about effectiveness. Still, information about learning and skills is quite useful.

3.5 Types of Research Methodology

The classification of research can be based on the discipline as the case with educational research and scientific research. Educational research is methods used to verify knowledge that will help in solving educational problems or assist the educationists achieve their goals. Most researches can be classified into one of the following broad categories or types: historical research, descriptive research, correlation research and experimental research.

3.5.1 Historical Research

The historical research deals with the determination, evaluation and explanation of the past events for the purpose of understanding the present and predicting the future. In historical research, one is looking at an event which has already taken place. The data is already there, it is just a matter of searching through, without generating new data. This makes people aware of what has happened in the past so that one can learn from the past failure and successes in preparation for the present and future. Since many records are not adequately kept or are incomplete and often lost, historical method of research is inconclusive. An example of a historical research topic is "Impact of Western Education in Nigeria". There are two sources of historical data:

- Primary source
- Secondary source

Primary Source is information obtained from first-hand; observed or available within one's immediate environment. Examples include stories told by the head of the community and chiefs.

Secondary Source deals with bibliographies, references and documents recorded by someone else. Generally, in this type of research statistical analysis is rarely used.

3.5.2 Descriptive Research

This is aimed at describing the characteristics of subject. Much of the early work in science was descriptive, since it was necessary to know something about the characteristic of the subjects before trying to study more complex research questions. There is still a great deal to be known about the students and teachers who are the usual subjects in educational research.

3.5.3 Correlation Research

In a correlation research, the researcher selects a sample and members of the selected group are measured on the variable being studied. Therefore, a correlation is computed between scores by all members of the group on the two variables. Correlation studies can be classified as relationship studies and predictions studies: while the relationship studies are concerned with the relationship between measures of different variables obtained at approximately the same time to understand the factors that make up complex characteristics, the prediction studies are concerned with measuring a variable that can be used to predict some future event.

3.5.4 Experimental Research

Most of the basic experimental designs used in education and behavioural sciences have been adapted from the physical and biological sciences. Experimental research methods are sometimes difficult to apply to solve certain educational problems. The typical experimental design in education involves the selection of a sample of subjects, random division of these subjects to two groups, the experimental and control group; the exposure of the experimental group to a treatment called the independent variable, which is withheld from the control group; and the evaluation of the groups on the dependent variable, the variable to be changed. There are two essential characteristics of an experimental method: during research designs, there are treatments that manipulate the dependent variable; and the subjects who participate are randomly assigned to the treatments.

3.5.5 Quasi-experimental Research

In this research method, subjects may not be assigned randomly and in some cases certain quasi-experimental designs collected from this can only be accepted with less confidence.

SELF ASSESSMENT EXERCISE

What do you consider to be the major limitations of historical research? Discuss the criteria for evaluating a historical research.

4.0 CONCLUSION

Research is a process of inquiry, investigation, close scrutiny and discovery. Every time you seek an answer to a question, you are undertaking a research, however, small. Though research may seem tedious, it can be made easy, through careful planning and design. In this unit, we look at some problems a researcher is likely to encounter, how to choose a research topic, motivation for research. A researcher must decide in which area of time his problem is mainly centred. Based on that decision he selects his basic research method, the distinctive criteria of which he must carefully observe in the investigation of his problem.

5.0 SUMMARY

In this unit, we have considered the basic concepts related to research**;** such as types of research, research processes, hypothesis, research methods and research funding. We also looked at the considerations to design your research approach, overall goal in selecting research method.

6.0 TUTOR-MARKED ASSIGNMENT

1. Explain the concept of research.
2. What do you understand by a historical research?
3. What are the purposes of historical research?

7.0 REFERENCES/FURTHER READING

Osuala, E. C. (1982). *Introduction to Research Methodology*. Awka Rd Onitsha, Nigeria: Africana-Fep Publisher Limited.

Okoro, E. (2002). *Quantitative Techniques in Urban Analysis.* Ibadan: Kraft Books Ltd.

Kerlinger, Fred N. (1964). *Foundations of Behavioural Research*. New York: Holt, Rinehart and Winton.

Whitney, F.L. (1968). *The Elements of Research.* New York: Prentice-Hall.

UNIT 3 BASIC RESEARCH METHODOLOGY (II)

CONTENTS

1.0 INTRODUCTION

Research is a process of inquiry, investigation, close scrutiny and discovery. Every time you seek an answer to a question, you are undertaking a piece of research, however small. Though research may seem tedious, it can be made easy, through careful planning and design. In this unit we will look at some problems a researcher is likely to encounter, how to choose a research topic, motivation for research among other issues.

2.0 OBJECTIVES

In this unit, we will discuss the different types of techniques of data collection, reliability of instrument, how to select suitable research topics and others.

3.0 MAIN CONTENT

3.1 Instruments of Data Collection

There are different techniques or tools used in collecting information in research. The research findings of any study are no more accurate than the measures or tools by which these findings are obtained. In the behavioural sciences and education, a very large number of human phenomena must be measured if they are to be studied. Researchers have developed a variety of procedures for measuring human characteristics and behaviour. The measuring devices developed for collecting a data are called *Research Instruments*. Several R*esearch Instruments* are available for collecting data for a research work. These include tests, and pencil tests, questionnaires, interviews, socio-metric techniques, and anecdotal records. The most commonly used in education research is questionnaires and interviews.

3.2 Questionnaires Approach

This is a method of obtaining information through the answers supplied by a respondent who fills the materials forwarded to him or her. Questionnaires usually contain questions aimed at getting specific information on a variety of topics. Usually, there are no right or wrong answers. Questions may be of either the closed form (structured) in which the question permits only certain response (as the case in multiple choice), or the open form (unstructured), in which the subject makes any response he wishes. It takes the form of fill-the-gap, tick-the-appropriate-column or give-your-opinion. It can also take the form of strongly Agree to strongly Disagree. The entire work can be divided into sections from personal data to answering specific questions to answer the research questions and hypothesis.

3.2.1 Advantages and Disadvantages of Questionnaires

Advantages

1. It is economical since it could be sent through mail.
2. More people are reached within a short period.
3. Respondents are likely to give honest opinion because the researcher is not present (most especially when the questionnaire is mailed).
4. It is possible to be administered by another person other than the researcher.

Disadvantages

1. It is very difficult to check the motivation of the respondents
2. It is very costly to prepare questionnaire and the materials to be used.
3. Clarification of points, in case of ambiguity, may not be possible because of the absence of the researcher when the questionnaire is mailed.
4. The number of returns may be few, especially when there is no follow up.

3.2.2 Guidelines for Designing a Questionnaire

- Include brief and clear instructions printed in bold letters.
- Focus your questionnaire items on the research hypothesis or research question.
- The first set of items is to be attention catching and not controversial. It should be clearly worded.
- Let the items be from simple to complex and organised in sequence.
- Avoid questions, which demand two separate answers.
- Short items are preferable to long ones.
- Avoid biased or misleading questions.
- Use the level of language of the respondents.
- It is advisable to do a follow-up to get high returns of the questionnaires.

3.3 Interview Approach

It is a method in descriptive research such as surveys to collect information from others in a face-to-face contact. That is, data are collected through direct verbal interaction between two individuals; usually the communication is two-way. It must be noted that interview could be conducted through the media such as telephone and there is no face-to-face contact. However, there is two-way communication. Through an interview technique, the researcher gets more information, which would not have been possible with other methods, and the respondents can ask for clarification at any stage of the interview.

3.3.1 Types of Interview

- **Structured Interview**

This is rigidly standardized and very formal. The same questions in the same order are presented to various respondents who are restricted to a predetermined list of alternative answers. Although it does not allow in-depth probing; it is scientific in approach.

- **Unstructured Interview**

In this case, the respondents are free to express their opinion or feelings. There is flexibility with few restrictions. The amount of time and the questions asked vary from one respondent to another. However, it is rarely used for testing hypothesis because subjects' response cannot be compared.

3.3.2 Advantages and Disadvantages of Interview Approach

Advantages

1. It is flexible and applicable to solving different types of research problems.
2. The researcher may get more information from the subjects and know more than he has prepared.
3. There is room for better understanding of the questions because there is two-way communication.
4. It is easily applicable to people of different age groups and educational background. That is, it takes care of children and the illiterates who cannot read and write.
5. It permits an interviewer to study nonverbal cues. What the respondents say and how it is said can both be observed

Disadvantages

1. Time consuming to conduct interview for large group of people.
2. The respondents may not give honest opinion because of the presence of the researcher or because he/she is trying to please the interviewer.
3. The socio-economic status and sex of the interviewer may affect the respondents.
4. It is often difficult to generalise from unstructured interview and at times difficult to quantify the information gathered for analysis.

3.4 Observation

Observation can be viewed as a process where individuals or a group of people are commissioned to watch and record the happenings or events, or even study behaviour patterns in settings of interest. Observation could be direct or differential. Direct observation is a direct means of collecting the information required by the research i.e. the researcher makes use of another person's observation.

3.4.1 Advantages and Disadvantages of Observation

Advantages

1. It is better than other methods as the research would have a direct contact with phenomenon under study.
2. It provides very reliable and valid measure of the variables being observed.

Disadvantages

1. The recording instrument is always expensive and very delicate to use by untrained personnel, e.g. video, tape/camera.
2. No two people have seen an event in the same way.

3.5 Reliability of Instrument

It is important to ascertain the worth of the instrument a researcher is interested in before he begins to use it. The quality of the instrument used in research is very important for the conclusions the researcher draws based on the information they obtain using these instruments. The quality is determined by its validity and reliability.

3.6 Methods of Administering Instruments

Several methods could be used to collect data in educational researches. There is, however, some common method. These are:

3.6.1 Direct Administration

Here, the researcher can directly administer to respondents and wait for the respondents to complete and return them .e.g. questionnaire, interview, observation, etc. Most of the experimental studies require the researcher to physically administer the instrument.

3.6.2 Postal Administration

Some instruments could be administered by post. For example, questionnaire when filled can be returned by postage. The disadvantage in this method of administering instrument includes unreliability of the postal agency in this country. Also, respondents may not be quick or fast enough to return the questionnaires. If not, it is uncertain that high returns of the instrument can be achieved.

3.6.3 Telephone Method

This is the administering of the instrument through telephone. The items on the questionnaire are usually obtained immediately where respondents are willing to cooperate. This method could be efficient where telephone services are efficient and not too expensive.

3.7 Selecting a Suitable Research Topic

Selecting a research topic is a crucial step in any research effort. It is very important that a researcher knows how to identify a researchable problem. According to Dewey, 'a problem arises out of some felt difficulties' *Sax, 1979*. One of the reasons why student researchers have difficulty in selecting a research problem is that they often get confused about what a researchable problem could be. What then is a problem? Good (1973) defined problem as a perplexing situation translated into a question or series of questions that help to determine the direction of subsequent inquiry. There are times when problems are identified from dissatisfaction or worry: that is, a problem materialises when a researcher senses that something is not right and needs investigation or further explanation. Note that researchable problems could be obtained. These could be obtained in journals, research reports, dissertation abstracts, education index, and student research projects (dissertation or theses). (See *O.A Opadokun)*

4.0 CONCLUSION

It has always been of interest to people on how to cope with their environment and understand the nature of the phenomena the environment presents to their senses. The means by which they strive to achieve these goals may be broadly categoriesd as: experience, reasoning and research.

5.0 SUMMARY

In this unit, we have looked at factors to consider in designing your research approach, overall goal in selecting research methods, four levels of research results and the different types of research methodology. Techniques of data collection were also discussed.

6.0 TUTOR-MARKED ASSIGNMENT

1. Discuss the various types of Research
2. What are the advantages of interview approach?
3. Discuss the importance of reliability of instrument in research methodology.

7.0 REFERENCES/FURTHER READING

Osuala, E.C. (1982). *Introduction to Research Methodology*. Awka Rd Onitsha, Nigeria: Africana-Fep Publisher Limited.

Okoro, E. (2002) *Quantitative Techniques in Urban Analysis.* Ibadan: Kraft Books Ltd.

Kerlinger, Fred N. (1964). *Foundations of Behavioural Research*. New York: Holt, Rinehart and Winton.

Whitney, F.L. (1968).*The Elements of Research*. New York: Prentice-Hall.

UNIT 4 NATURE OF SCIENCE

CONTENTS

1.0 INTRODUCTION

Understanding how science works allows one to easily distinguish science from non-science. Thus, to understand biological evolution, or any other science, it is essential to begin with the nature of science. Over the course of human history, people have developed many ideas about the physical, biological, psychological, and social worlds, which are interconnected and validated. Successive generations have relied upon those ideas to achieve an increasingly comprehensive and reliable understanding of the human race and its environment. The means used to develop these ideas are particular ways of observing, thinking, experimenting, and validating. These ways represent a fundamental aspect of the nature of science and reflect how science tends to differ from other modes of knowledge.

2.0 OBJECTIVES

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

- define the term science
- distinguish between science and other fields of human endeavour
- explain the nature of science
- enumerate the importance of science
- identify personal characteristics of a scientist.

3.0 MAIN CONTENT

3.1 What is Science?

Science is an extension of the human being's inherent quest for knowledge and understanding, which is part of us from birth. It is a particular way of understanding the natural world, which allows us to connect the past with the present and both with the future. Science is premised upon the belief that our senses, and extensions of those senses through the use of instruments, can give us accurate information about the universe. Science follows very specific "rules" and its results are always subject to testing and revision, if necessary. These characteristics of science might seem like constraints to creativity and imagination; but in fact, science does not exclude, but often benefits from these aspects of life. The scientific method of enquiry offers an objective way to evaluate the truth or otherwise of information. The late astronomer Carl Sagan, Ph.D., has pointed out that "Science is a way of thinking much more than it is a body of facts."

3.1.1 Understanding the Meaning of Science

In scientific terms, "theory" does not mean "guess" or "hunch" as it does in everyday usage. Scientific theories are explanations of natural phenomena built up logically from testable observations and hypotheses. Scientists most often use the word "fact" to describe an observation, but "fact" can also be employed to refer to something that has been tested or observed so many times that there is no longer a compelling reason to keep testing or looking for examples. Usually "faith" refers to beliefs that are accepted without empirical [observed] evidence. Most religions have tenets of faith. Science differs from religion because it is the nature of science to accept or believe only those explanations, which have been tested and re-tested against the natural world. Thus, scientific explanations are likely to be built on and modified with new information and new ways of looking at old information. This is quite different from most religious beliefs. From the forgoing, we can say that “belief" is really not an appropriate term to use in science: no knowledge is accepted in science unless it has been tested. Testing is a fundamental path to knowing. If there is any component of faith to science, it is the assumption that the universe operates according to regularities. This "faith" is very different from religious faith.

3.1.2 Theory as a Major Component of Science

A theory is an attempt to synthesise and integrate, empirical data for maximum clarification and unification. Every individual has a number of personal theories, which he/she comes up with by postulations, and assumptions, which are true in varying degrees. The individual relies on these postulations and assumptions to varying degrees when he/she makes deductions and decisions. The schoolteacher, for instance, has many theories about education. These may be based partially on personal experience, or on his reading of relevant literature. It can be based on the teacher’s personal philosophy. The purposes served by theory can be summarised as follows:

1) Theory combines isolated bits of empirical data into a broader conceptual scheme that has wider applicability and predictability. With a theory, the investigator obtains a deeper understanding of data and enables the translation of empirical findings into a form, which is more readily understood, retained and adaptable.
2) Theory permits the prediction of the occurrence of phenomena and enables the investigator to postulate and, eventually, to discover hitherto unknown and unobserved phenomena.
3) Theory acts as a guide to discovery of facts; it identifies crucial aspects to be investigated and crucial questions to be answered. By so doing, it stimulates research into areas of knowledge that are yet to be explored.
4) Theory is based on the assumption that detailed empirical findings are exceptional cases arising from laws that are more general. Progress cannot be made as long as observations are simply accumulated. Theories cannot develop without experimental facts, anything less will produce grossly inadequate or incorrect theories. For example, the progress of psychiatry as a science will continue to be limited as long as the insane are viewed as possessed by a devil. Just as facts underlie theories, theories underlie facts, each raising the other on a spiral to evermore precise scientific formulations. Research and theory go hand in hand; theory guides and stimulates research while

research tests and stimulates development of theory, resulting in more adequate theories and better and clearer facts.

3.1.3 Composition of Science

Science, Technology and Mathematics, make up a successful scientific endeavour. Although each of these human enterprises has a character and history of its own, each is dependent on and reinforces the others. From an organisational point of view, Science can be regarded as the collection of all of the different scientific fields, or content disciplines. From anthropology to zoology, there are dozens of such disciplines. They differ from one another in many ways, including history, phenomena studied, techniques and language used, and kinds of outcomes desired. With respect to purpose and philosophy, however, all are equally scientific and together make up the same scientific endeavour. The advantage of having disciplines is that they provide a conceptual structure for organising research and research findings.

The disadvantage is that their divisions do not necessarily match the way the world works, and they can make communication difficult. In any case, scientific disciplines do not have fixed borders. Physics shades into chemistry, astronomy, and geology, as does chemistry into biology and psychology, and so on. New scientific disciplines (astrophysics and sociobiology, for instance) are continually being formed at the boundaries of others. Some disciplines grow and break into sub disciplines, which then become disciplines in their own right.

Universities, industries, and governments are also part of the structure of the scientific endeavour. University research usually emphasises knowledge for its own sake, although much of it is also directed toward practical problems. Universities, of course, are also particularly committed to educating successive generations of scientists. Industries and businesses usually emphasise research directed to practical ends, but many also sponsor research that has no immediately obvious applications, partly on the premise that it will be applied fruitfully in the long run.

3.2 Characteristics of Science

1. Science is a truth-seeking process. It is not a collection of unassailable "truths." It is, however, a self -correcting discipline. Such corrections may take a long time. For example, the medical practice of bloodletting went on for centuries before its futility was realised, but as scientific knowledge accumulates the chance of making substantial errors decreases.

2. Certainty is elusive in science, and it is often hard to give categorical "Yes" or "No" answers to scientific questions. To determine whether bottled water is preferable to tap water, for example, one would have to design a lifelong study of two large groups of people whose lifestyles were similar in all respects except for the type of water they consume. This is virtually impossible. We therefore have to rely on less-direct evidence in formulating many of our conclusions.

3. It may not be possible to predict all consequences of an action, no matter how much advanced research has been done. When chlorofluorocarbons (CFCs) were introduced as refrigerants, no one could have predicted that 30 years later they would have an impact on the ozone layer. If something undesirable happens, it is not necessarily because someone has been negligent.

4. Any new finding should be examined with scepticism. Healthy scepticism does not mean unwillingness to believe. Sceptics base their beliefs on scientific proof and do not swallow information uncritically.

5. No major lifestyle change should be based on any one study. Others should independently confirm results. Keep in mind that science does not proceed by "miracle breakthroughs" or "giant leaps." It plods along, taking many small steps, slowly building towards a consensus.

6. Studies have to be carefully interpreted by experts in the field. An association of two variables does not necessarily imply cause and effect. As an extreme example, consider the strong association between breast cancer and the wearing of skirts. Obviously, wearing skirts does not cause the disease. Scientists, however, sometimes show an amazing aptitude for coming up with inappropriate rationalisations for their pet theories.

7. Repeating a false notion does not make it true. Many people are convinced that sugar causes hyperactivity in children -- not because they have examined studies to this effect but because they have heard that it is so. In fact, a slate of studies has demonstrated that, if anything, sugar has a calming effect on children.

8. There often are legitimate opposing views on scientific issues. But it is incorrect to conclude that science cannot be trusted because for every study there is an equal and opposite study. It is always important to take into account who carried out a given

study, how well it was designed, and whether anyone stands to gain financially from the results. Be mindful of whom the "they" is in "they said that . . . ." In many cases, what “they say" is only gossip, inaccurately reported.

3.3 The Scientists’ Beliefs

Scientists share certain basic beliefs and attitudes about what they do and how they view their work. These have to do with the nature of the world and what can be learned about it.

3.3.1 Possibility of Understanding the Complexity of the World

Scientists believe that the things and events in the universe occur in consistent patterns that are comprehensible through careful, systematic study. Scientists believe that with the intellect, and with the aid of instruments that extend the senses, people can discover patterns in all nature. Science also assumes that the universe is, as its name implies, a vast single system in which the basic rules are the same everywhere. Knowledge gained from studying one part of the universe is applicable to other parts. For instance, the same principles of motion and gravitation that explain the motion of falling objects on the surface of the earth also explain the motion of the moon and the planets. With some modifications over the years, the same principles of motion have applied to other forces and to the motion of everything, from the smallest nuclear particles to the most massive stars, from sailboats to space vehicles, from bullets to light rays.

3.3.2 Scientific Ideas are not static

Science is a process of producing knowledge. The process depends on making careful observations of phenomena, and on inventing theories for making sense out of those observations. Change in knowledge is inevitable because new observations challenge prevailing theories. No matter how well one theory explains a set of observations, it is possible that another theory may fit just as well or better, or may fit a still wider range of observations. In science, the testing and improving and occasional discarding of theories, whether new or old, go on all the time. Scientists assume that even if there is no way to secure complete and absolute truth, increasingly accurate approximations can be made to account for the world and how it works.

3.3.3 Knowledge of Scientific Research is Durable

Most scientific knowledge is durable. This is despite the fact that scientists reject the notion of attaining absolute truth and accept that it is only part of nature to have a level of uncertainty. Science preserves a tradition of modification of ideas, rather than their outright rejection. Ideas that are greatly criticized tend to survive and grow more precise until they become widely accepted. For example, in formulating the theory of relativity, Albert Einstein did not discard the Newtonian laws of motion but rather showed them to be only an approximation of limited application within a more general concept. (The National Aeronautics and Space Administration use Newtonian mechanics, for instance, in calculating satellite trajectories). Moreover, the growing ability of scientists to make accurate predictions about natural phenomena provides convincing evidence that progress is being made in man's understanding of how the world works. Continuity and stability are as characteristic of science as change is, and confidence is as prevalent as tentativeness.

3.3.4 The Limitations of Science

Many matters cannot usefully be examined in a scientific way. There are, for instance, beliefs that by their very nature cannot be proved or disproved; for example, the existence of supernatural powers and beings, or the true purpose of life. In other cases, a scientific approach that may be valid is likely to be rejected as irrelevant by people who hold to certain beliefs (such as in miracles, fortune telling, astrology, and superstition). Nor do scientists have the means to settle issues concerning good and evil, although they can sometimes contribute to the discussion of such issues by identifying the likely consequences of particular actions, which may be helpful in weighing alternatives.

3.4 Scientific Inquiry

Fundamentally, the various scientific disciplines are alike in their reliance on evidence, the use of hypothesis and theories, the kinds of logic used, and much more. Nevertheless, scientists differ greatly from one another in what phenomena they investigate and in how they go about their work; in the reliance they place on historical data or on experimental findings and on qualitative or quantitative methods; in their recourse to fundamental principles; and in how much they draw on the findings of other sciences. Still, the exchange of techniques, information, and concepts goes on all the time among scientists, and there are common understandings among them about what constitutes an investigation that is scientifically valid.

Scientific inquiry is not easily described apart from the context of particular investigations. Simply, there is no fixed set of steps that scientists always follow, no single path that leads them unerringly to scientific knowledge. There are, however, certain features of science that give it a distinctive character as a mode of inquiry. Although those features are especially characteristic of the work of professional scientists, everyone can exercise them in thinking scientifically about many matters of interest in everyday life.

3.4.1 Science Demands Evidence

Eventually, the validity of scientific claims is established by referring to observations of phenomena. Hence, scientists concentrate on getting accurate data. Such evidence is obtained by observations and measurements taken in situations that range from natural settings (such as a forest) to completely contrived ones (such as the laboratory). To make their observations, scientists use their own senses, instruments (such as microscopes) that enhance those senses and instruments that tap characteristics quite different from what humans can sense (such as magnetic fields). Scientists observe passively (earthquakes, bird migrations), make collections (rocks, shells), and actively probe the world (as by boring into the earth's crust or administering experimental medicines).

In some circumstances, scientists can control conditions deliberately and precisely to obtain their evidence. They may, for example, control the temperature, change the concentration of chemicals, or choose which organisms mate with which others. By varying just one condition at a time, they can hope to identify its exclusive effects on what happens, uncomplicated by changes in other conditions.

Often, however, control of conditions may be impractical (as in studying stars), or unethical (as in studying people), or likely to distort the natural phenomena (as in studying wild animals in captivity). In such cases, observations have to be made over a sufficiently wide range of naturally occurring conditions to infer what the influence of various factors might be. Because of this reliance on evidence, great value is placed on the development of better instruments and techniques of observation, and the findings of any one investigator or group are usually checked by others.

3.4.2 Science is a Blend of Logic and Imagination

Although all sorts of imagination and thought may be used in coming up with hypotheses and theories, sooner or later scientific arguments must conform to the principles of logical reasoning, that is, to testing the validity of arguments by applying certain criteria of inference, demonstration, and common sense. Scientists may often disagree about the value of a particular piece of evidence, or about the appropriateness of particular assumptions that are made - and therefore disagree about what conclusions are justified. Nevertheless, they tend to agree about the principles of logical reasoning that connect evidence and assumptions with conclusions.

Scientists do not work only with data and well-developed theories. Often, they have only tentative hypotheses about the way things may be. Such hypotheses are widely used in science for choosing what data to pay attention to and what additional data to seek, and for guiding the interpretation of data. In fact, the process of formulating and testing hypotheses is one of the core activities of scientists. To be useful, a hypothesis should suggest what evidence would support it and what evidence would refute it. A hypothesis that cannot, in principle, be put to the test of evidence may be interesting, but it is not likely to be scientifically useful.

The logical and meticulous examination of evidence is necessary; it is not by itself, sufficient for the advancement of science. Scientific concepts do not emerge automatically from data or from any amount of analysis alone. Inventing hypotheses or theories to imagine how the world works and then figuring out how they can be put to the test of reality is as creative as writing poetry, composing music, or designing skyscrapers. Sometimes discoveries in science are made unexpectedly, even by accident. However, knowledge and creative insight are usually required to recognise the meaning of the unexpected. Aspects of data that have been ignored by one scientist may lead to new discoveries by another.

3.4.3 Science Explains and Predicts

Scientists strive to make sense of observations of phenomena by constructing explanations for them that use, or are consistent with, currently accepted scientific principles. Such explanatory theories may be either sweeping or restricted, but they must be logically sound, incorporating a significant body of scientifically valid observations. The credibility of scientific theories often comes from their ability to show relationships between phenomena that previously seemed unrelated. The theory of moving continents, for example, has grown in credibility as it has shown relationships among such diverse phenomena as earthquakes, volcanoes, and the match between types of fossils on different continents, the shapes of continents, and the contours of the ocean floors.

The essence of science is validation by observation. But it is not enough for scientific theories to fit only the observations that are already known.

Theories should also fit additional observations that were not used in formulating the theories in the first place; that is, theories should have predictive power. Demonstrating the predictive power of a theory does not necessarily require the prediction of events in the future. The predictions may be about evidence from the past that has not yet been found or studied. A theory about the origins of human beings, for example, can be tested by new discoveries of human-like fossil remains. This approach is clearly necessary for reconstructing the events in the history of the earth or of the life forms on it. It is also necessary for the study of processes that usually occur very slowly, such as the building of mountains or the aging of stars. Stars, for example, evolve more slowly than we can usually observe. Theories of the evolution of stars, however, may predict unsuspected relationships between features of starlight that can then be sought in existing collections of data about stars.

3.4.4 Scientists Try to Identify and Avoid Bias

When faced with a claim that something is true, scientists respond by asking what evidence supports it. But scientific evidence can be biased about how the data are interpreted, recorded or reported, or even in the choice of what data to consider in the first place. Scientists' nationality, sex, ethnic origin, age, political convictions, and so on may incline them to look for or emphasise one or another kind of evidence or interpretation. For example, for many years the study of primates by male scientists focused on the competitive social behaviour of males. It was not until female scientists entered the field that the importance of female primates' community-building behaviour was recognised. Bias attributable to the investigator, the sample, the method, or the instrument may not be completely avoidable in every instance, but scientists want to know the possible sources of bias and how bias is likely to influence evidence. Scientists are expected, to be as alert to possible bias in their own work as in that of other scientists, although such objectivity is not always achieved. One safeguard against undetected bias in an area of study is to have many different investigators or groups of investigators working in it.

3.4.5 Science is not authoritarian

It is appropriate in science, as elsewhere, to turn to knowledgeable sources of information and opinion, usually people who specialise in relevant disciplines. But esteemed authorities have been wrong many times in the history of science. In the end, no scientist, however famous or highly placed, is empowered to decide for other scientists what is true, for none are believed by other scientists to have special access to the truth. There are no pre-established conclusions that scientists must reach based on their investigations. It is undeniable that new ideas that do not mesh well with mainstream ideas may encounter vigorous criticism, and scientists investigating such ideas may have difficulty obtaining support for their research. In reality, challenges to new ideas are the legitimate business of science in building valid knowledge. Even the most prestigious scientists have occasionally refused to accept new theories despite the existence of enough accumulated evidence to convince others. In the long run, however, theories are judged by their results: When someone comes up with a new or improved version that explains more phenomena or answers more important questions than the previous version, the new one eventually takes its place.

3.4.6 Science is a Complex Social Activity

Scientific work involves many individuals doing many different kinds of work and goes on to some degree in all nations of the world. Science and its application involve the participation of men and women of all ethnic and national backgrounds. These people-scientists and engineers, mathematicians, physicians, technicians, computer programmers, librarians, and others may focus on scientific knowledge either for its own sake or for a particular practical purpose, and they may be concerned with data gathering, theory building, instrument building, or communicating.

As a social activity, science inevitably reflects social values and viewpoints of scientists. The history of economic theory, for example, has paralleled the development of ideas of social justice at one time, economists considered the optimum wage for workers to be no more than what would just barely allow the workers to survive. Before the twentieth century, and well into it, women and people of color were essentially excluded from most of science by restrictions on their education and employment opportunities; the remarkable few who overcame those obstacles were even then likely to have their work belittled by the science establishment.

The direction of scientific research is affected by informal influences within the culture of science itself, such as prevailing opinion on what questions are most interesting or what methods of investigation are most likely to be fruitful. Elaborate processes involving scientists themselves have been developed to decide which research proposals receive funding, and committees of scientists regularly review progress in various disciplines to recommend general priorities for funding. Science goes on in many different settings. Scientists are employed by universities, hospitals, business and industry, government, independent research organisations, and scientific associations. They may work alone, in small groups, or as members of large research teams. Their places of work include classrooms, offices, laboratories, and natural field settings from space to the bottom of the sea.

Because of the social nature of science, the dissemination of scientific information is crucial to its progress. Some scientists present their findings and theories in papers that are delivered at meetings or published in scientific journals. Those papers afford the scientists a level of exposure of their works to others. It also exposes their ideas to criticism by other scientists, and, of course, keeps the science community abreast of scientific developments in other places around the world. The advancement of information science (knowledge of the nature of information and its manipulation) and the development of information technologies (especially computer systems) affect all sciences. Those technologies speed up data collection, compilation, and analysis; make new kinds of analysis practical; and shorten the time between discovery and application.

3.5 Ethical Principles in the Conduct of Science

There are some basic principles guiding the conduct and practice of science; we will look at some of these principles in this section for better understanding of the nature of science.

3.5.1 Accuracy, Good Record Keeping, Openness and Replication

Scientists are expected to conduct themselves according to the ethical norms of science. Certain deeply entrenched traditions such as accurate recordkeeping, openness, and replication, as well as the critical review of one's work by peers, serve to keep the vast majority of scientists well within the bounds of ethical professional behaviour.

Sometimes, however, the pressure to get credit for being the first to publish an idea or observation leads some scientists to withhold information or even to falsify their findings. Such a violation of the very nature of science impedes science. When discovered, it is strongly condemned by the scientific community and the agencies that fund research.

3.5.2 Protection of Lives and Properties

Another important aspect of scientific ethics relates to possible harm that could result from scientific experiments. One aspect is the treatment of live experimental subjects. Modern scientific ethics require that due regard must be given to the health, comfort, and well-being of animal subjects. Moreover, research involving human subjects may be conducted only with the informed consent of the subjects, even if this constraint limits some kinds of potentially important research or influences the results. Informed consent entails full disclosure of the risks and intended benefits of the research and the right to refuse to participate. In addition, scientists must not knowingly subject co-workers, students, the neighbourhood, or the community to health or property risks without their knowledge and consent.

3.5.3 Consideration for Possible Harmful Effects of the Application of Scientific Findings

The ethics of science also relates to the possible harmful effects of applying the results of research. The long-term effects of science may be unpredictable, but some idea of what applications are expected from scientific work can be ascertained by knowing who is interested in funding it. If, for example, the Department of Defense offers contracts for working on a line of theoretical mathematics, mathematicians may infer that it has application to new military technology and therefore would likely be subject to secrecy measures. Military or industrial secrecy is acceptable to some scientists but not to others. Whether a scientist chooses to work on research of great potential risk to humanity, such as nuclear weapons or germ warfare, is considered by many scientists to be a matter of personal ethics, not one of professional ethics.

SELF ASSESSMENT EXERCISE

1. Give a comprehensive definition of science?
2. What is the relevance of science to our economy?

4.0 CONCLUSION

Science is a truth-seeking process. It is not a collection of unassailable "truths". It is, however, a self -correcting discipline. Such corrections may take a long time. Scientists can bring information, insights, and analytical skills to bear on matters of public concern. Often they can help the public and its representatives to understand the likely causes of events such as natural and technological disasters and to estimate the possible effects of projected policies such as ecological effects of various farming methods. Often they can testify to what is not possible. In playing this advisory role, scientists are expected to be especially careful in trying to distinguish fact from interpretation, and research findings from speculation and opinion; that is, they are expected to make full use of the principles of scientific inquiry.

5.0 SUMMARY

In this unit, we have closely examined the following; the definition of science, the significance of theory to science, composition of science, characteristics of science, the Scientists Beliefs, scientific Inquiry, and ethical principles in the conduct of science.

6.0 TUTOR - MARKED ASSIGNMENT

1. Theory serves some purposes in research. Briefly mention a few of them.
2. What is the purpose of science?
3. Enumerate some of the basic ethics of a scientist?

7.0 REFERENCES/FURTHER READING

E.C. Osuala, (1982). *Introduction to Research Methodology*. Awka Rd Onitsha, Nigeria: Africana-Fep Publisher Limited.

E. Okoro (2002). *Quantitative Techniques in Urban Analysis.* Ibadan: Kraft Books Ltd.

Whitney, F.L. (1968). *The Elements of Research*: New York: Prentice-Hall.

Kerlinger, Fred N. (1964) *Foundations of Behavioural Research*. New York: Holt, Rinehart and Winton.

UNIT 5 SOME BASIC CONCEPTS IN SOCIAL STATISTICS

CONTENTS

1.0 INTRODUCTION

Statistical tools are essential for social scientists. Basic concepts of statistics, especially randomness and averaging, provide the foundations for measuring concepts, designing studies, estimating quantities of interest, and testing theories and conjectures. Social research refers to research conducted by social scientists (primarily within sociology and social psychology), but also within other disciplines such as social policy, human geography, political science, social anthropology and education. Sociologists and other social scientists study diverse things: from census data on hundreds of thousands of human beings, through the in-depth analysis of the life of a single important person to monitoring what is happening on a street today - or what was happening a few hundred years ago.

2.0 OBJECTIVES

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

- define the meaning of social statistics
- explain the purpose of social statistics
- enumerate the methods used in social statistics
- explain the aims, ethics of social research etc.

3.0 MAIN CONTENT

3.1 Definition of Social Statistics

Social statistics is the use of statistical measurement systems to study human behaviour in a social environment. This can be accomplished through polling a particular group of people, evaluating a particular subset of data obtained about a group of people, or by observation and statistical analysis of a set of data that relates to people and their behaviours.

3.1.1 Purpose of Social Statistics

Social scientists use social statistics for many purposes, including the following, though the list is not exhaustive:

1) The evaluation of the quality of services available to a particular group or organisation,
2) Analysing behaviours of groups of people in their environment and special situations,
3) Determining the wants or needs of people through statistical sampling.

3.2 Methods Used in Social Statistics

Social scientists use many different methods in order to describe, explore and understand social life. Social methods can generally be subdivided into two broad categories.

3.2.1 Quantitative Methods

These are those techniques used in social statistics to quantify social phenomena and collect and analyse numerical data. With this method, smaller number of attributes is considered in a wide variety of cases with the focus being to establish links between all these attributes. Researchers commonly employ surveys and questionnaires as tools of quantitative research. In addition, statistical data (for example, censuses or the results of social attitudes surveys) that has been gathered for other purposes may also be subjected to secondary analysis; this is another tool used under the quantitative method of social statistics.

3.2.2 Qualitative Methods

This method pays more attention to interpretations and comprehension of personal experiences in order to be able to understand the meaning of social phenomena. Less emphasis is placed on quantification. Thus, qualitative methods focus on finding the links among a larger number of attributes across relatively few cases. Commonly used qualitative methods include focus groups, participant observation, and other techniques. While very different in many aspects, both qualitative and quantitative approaches involve a systematic interaction between theories and data.

3.2.3 Similarities between Quantitative and Qualitative Methods

Researchers have come to realise that quantitative and qualitative methods can be complementary. Currently, researchers avoid any overemphasis on the significance of these differences between the two methods. They can be combined in a number of ways, for example:

1. Qualitative methods can be used in order to develop quantitative research tools. For example, focus groups could be used to explore an issue with a small number of people and the data gathered using this method could then be used to develop a quantitative survey questionnaire that could be administered to a far greater number of people allowing results to be generalised.

2. Qualitative methods can be used to explore and facilitate the interpretation of relationships between variables. For example researchers may inductively hypothesise that there would be a positive relationship between positive attitudes of sales staff and the amount of sales of a store. However, quantitative, deductive, structured observation of 576 convenience stores could reveal that this was not the case, and in order to understand why the relationship between the variables was negative the researchers may undertake qualitative case studies of four stores including participant observation.

3.3 The Ethics of Social Research

The primary assumptions of the ethics in social research are:

- Voluntary participation
- No harm to subjects
- Integrity

3.4 Aims of the Social Sciences

Having examined the assumptions of science, we are now in a position to address the question raised earlier: what does science have to offer to people who take an interest in society's problems? The ultimate goal of the social and all other sciences is to produce a cumulative body of verifiable knowledge. Such knowledge enables us to explain, predict, and understand the empirical phenomena that hinder us. Furthermore, a reliable body of knowledge could be used to improve the human condition. But what are scientific explanations? When can we make predictions? When are we justified in claiming that we understand empirical phenomena? The social scientist's aim is to improve general explanations for "Why?" questions. When scientists ask for an explanation of why a given event or behaviour has taken place, they ask for a systematic and empirical analysis of the antecedent factors that caused the event or behaviour.

The role of science is to establish general law that will govern the behaviour of empirical events or objects with which the science in question is concerned. The purpose of such governing laws is to enable us to connect together our knowledge of the separately known events, and to make reliable predictions of events as yet unknown, if science is in a highly developed state. These laws, which have been established, will form a hierarchy in which special laws appear as logical consequences of a small number of highly general laws. If the science is in an early stage of development, the laws may be merely the generalisations involved in ordering things into various classes.

3.5 Ordinary Human Inquiry

Before the advent of sociology and application of the scientific methods to social research, research was founded upon personal experiences, and received wisdom in the form of tradition and authority. Such approaches often led to errors such as inaccurate observations, over-generalisation, selective observations, subjectivity and lack of logic.

3.6 Foundations of Social Research

Social research (and social science in general) is based on logic and empirical observations. Social research involves the interaction between ideas and evidence. Ideas help social researchers make sense of evidence, and researchers use evidence to extend, revise and test ideas**.** Social research thus attempts to create or validate theories through data collection and data analysis, and its goal is exploration, description and explanation. It should never be mistaken with philosophy or belief. Social research focuses on finding the social patterns of regularity in social life. Usually it deals with social groups (aggregates of individuals), not individuals themselves (the science of psychology is an exception here). Research can also be divided into pure research and applied research. Pure research has no application on real life, whereas applied research attempts to influence the real world.

There are no laws in social science that parallel the laws in the natural science. A law in social science is a universal generalisation about a class of facts. A fact is an observed phenomenon, and observation means it has been seen, heard or otherwise experienced by researcher. A theory is a systematic explanation for the observations that relate to a particular aspect of social life. Concepts are the basic building blocks of theory and are abstract elements representing classes of phenomena. Axioms or postulates are basic assertions assumed to be true. Propositions are conclusions drawn about the relationships between concepts, based on analysis of axioms. Hypotheses are specified expectations about empirical reality, which are derived from propositions. Social research involves testing these hypotheses to see if they are true.

Social research involves creating a theory, operationalisation (measurement of variables) and observation (actual collection of data to test hypothesised relationship). Social theories are written in the language of variables, in other words, theories describe logical relationships between variables. Variables are logical sets of attributes, with people being the 'carriers' of those variables (for example, gender can be a variable with two attributes: male and female). Variables are also divided into *independent* variables (data) that influences the *dependent* variables (which scientists are trying to explain). For example, in a study of how different dosages of a drug are related to the severity of symptoms of a disease, a measure of the severity of the symptoms of the disease is a dependent variable and the administration of the drug in specified doses is the independent variable. Researchers will compare the different values of the dependent variable (severity of the symptoms) and attempt to draw conclusions.

3.7 Concept of Explanations in Social Science

Explanations in social theories can be idiographic (emphasising unique traits or functioning of individuals) or homothetic (relating to enactment of laws). An idiographic approach to an explanation is one where the scientists seek to exhaust the idiosyncratic causes of a particular condition or event, i.e. by trying to provide all possible explanations of a particular case. Sampling explanations tend to be more general with scientists trying to identify a few causal factors that impact a wide class of conditions or events. For example, when dealing with the problem of how people choose a job, idiographic explanation would be to list all possible reasons why a given person (or group) chooses a given job, while a sampling explanation would try to find factors that determine why job applicants in general choose a given job.

3.8 Types of Inquiry

Social research can be deductive or inductive. The inductive inquiry (also known as grounded research) is a model in which general principles (theories) are developed from specific observations. In deductive inquiry specific expectations of hypothesis are developed on the basis of general principles (i.e. social scientists start from an existing theory, and then search for proof). For example, in inductive research, if a scientist finds that some specific religious minorities tend to favour a specific political view, he may then extrapolate this to the hypothesis that all religious minorities tend to have the same political view. In deductive research, a scientist would start from a hypothesis that religious affiliation influenced political views and then begin observations to prove or disprove this hypothesis.

4.0 CONCLUSION

Statistics and statistical analyses have become key features of contemporary social science. Statistics is and has been perhaps most important in economics and psychology but is also employed in political science, sociology and anthropology. There is, however, currently a heated debate regarding the questionable uses and value of statistical methods in social science, especially in political science, with many statisticians questioning the policy conclusions of political partisans who overestimate the interpretive power that non-robust statistical methods such as simple and multiple linear regressions allow. Indeed, an important mantra that social scientists cite, but often forget, is that "correlation does not imply causation’’

5.0 SUMMARY

In this unit, we have looked at the definition of social statistics, purpose of social statistics, and methods used in social statistics. The ethics of social research, aims of the social sciences, ordinary human inquiry, and other issues related to social statistics were also discussed.

6.0 TUTOR-MARKED ASSIGNMENT

1. In your own words, define the term social statistic
2. What do you understand by methods of social statistics?
3. Is social statistics relevant to our political system?Discuss.

7.0 REFERENCES/FURTHER READING

Osuala, E.C. (1982). *Introduction to Research Methodology*. Awka Rd Onitsha, Nigeria: Africana-Fep Publisher Limited.

Okoro, E. (2002). *Quantitative Techniqes in Urban Analysis.* Ibadan: Kraft Books Ltd.

Kerlinger, Fred N. (1964). .*Foundations of Behavioural Research*. New York: Holt, Rinehart and Winton.

Whitney, F.L. (1968). *The Elements of Research*. New York: Prentice-Hall.