# NATIONAL OPEN UNIVERSITY OF NIGERIA

## SCHOOL OF EDUCATION

## COURSE CODE: BED 212:

## COURSE TITLE: FUNDERMENTALS OF DATA PROCESSING

# BED 212: FUNDERMENTALS OF DATA PROCESSING

## COURSE DEVELOPER & WRITER:

**DR. OPATEYE JOHNSON AYODELE**
**School of Education**
**National Open University of Nigeria**
**14/16 Ahmadu Bello Way, Victoria Island**
**Lagos. Nigeria.**

**COURSE CORDINATOR:**

**DR. INEGBEDION, JULIET O.**
**School of Education**
**National Open University of Nigeria**
**14/16 Ahmadu Bello Way, Victoria Island**
**Lagos. Nigeria.**

# INTRODUCTION

## BED 212: FUNDERMENTALS OF DATA PROCESSING

This course is designed to equip the students with knowledge of the data processing in technical and vocational education (TVE) especially Business Education research.

**WHAT YOU WILL LEARN**

You will learn the components of data processing, hardware and software components, file management and organization in Business Education- based data. It will also interest you to learn about basics of research, approaches and designs with data gathering techniques. Relevant statistical tools you can use to analyse your data would be across and how to write your research reports.

**COURSE AIMS**

This course aims at producing competent business educators who will be versed in organizational and research data processing in order to foster time and effective information that will be used in decision making. In order to enable you meet the above aims, modules constituting of units have been produced for your study. Apart from meeting the aims of the course as a whole, each course unit consists of learning objectives which are intended to ensure your learning effectiveness.

**COURSE OBJECTIVES**

The course objectives are meant to enable you achieve/acquire the following:

1) Gain in-depth knowledge of data processing and its functions in business organisations and educational settings.
2) Appreciate the historical perspective of data processing and how it evolves from manual to electronically based data processing.
3) Acquire skills in generating researchable problems and derive relevant research questions and hypotheses.
4) Understanding of the basic research designs to employ when conducting business education studies.
5) Appreciate various research data collection tools that are available to various business and educational projects
6) Demonstrate skills in utilizing appropriate statistical tools in analyzing research data and giving relevant interpretations to results generated.
7) Write comprehensive research reports after acquiring knowledge of the components of report writing.

**WORKING THROUGH THIS COURSE**

You are required to thoroughly work through all the units in this course. There are three modules in all with sixteen units.

**COURSE MATERIALS**

The major components of this course are

1. Course Guide

2. Study units
3. Text books
4. CDS
5. Practice In-text questions
6. Tutor Assignment file
7. E- Facilitation

**STUDY UNITS**

The breakdown of the three modules and study units are as follows: Every unit contains a list of references and further reading. Try as much as you can to get the listed books and further reading materials. You are also expected to approach the internet for further related reading materials. This is to widen as well as deepen the depth of understanding of this course

## MODULE ONE: INTRODUCTION TO DATA PROCESSING



## MODULE 2: DATA PROCESSING IN EDUCATIONAL RESEARCH



## MODULE 3: STATISTICS IN EDUCATIONAL RESEARCH

**PRESENTATION SCHEDULE**

The presentation schedule which is included in your course materials gives you the important dates for the completion of tutor-marked assignments and for attendance of tutorials. Remember, you are required to submit all your assignments on due dates. You should guard against falling behind in your work.

**ASSESSMENT**

Your assessment will be based on tutor-marked assignments (TMAs) and a final examination which you will write at the end of the course. You are required to complete four TMAs which will contain 20 multiple choice test items each.

**TUTOR MARKED ASSIGNMENTS (TMA)**

Every unit contains at least one or two assignments. You are advised to work through all the assignments and submit them for assessment. Your tutor will assess the assignments and select four, which will be marked and the best three will be selected which will constitute 30% of your final grade. The tutor-marked assignments may be presented to you in a separate file. Just know that for every unit there are some tutor-marked assignments for you. It is important you do them and submit for assessment.

**FINAL EXAMINATION AND GRADING**

At the end of the course, you will write a final examination which will constitute 70% of your final grade. In the examination which shall last for two hours, you will be required to answer 120 objectives questions in multiple choice, Yes/No and fill-in response format.

**COURSE MARKING SCHEME**

This table shows how the actual course marking is broken down.

| **Assessment** | |
|---|---|
| **Assignments** | Four assignments. Best three marks of the four count as 30% of course marks |
| **Final Examination** | 70% of overall course marks |
| **Total** | 100% of course marks |

# MODULE ONE: INTRODUCTION TO DATA PROCESSING

Unit 1: Concept of Data Processing

Unit 2: Historical Perspectives of Data Processing

Unit 3: Hardware Supports in Data Processing

Unit 4: Software Elements in Data Processing

Unit 5: Data Processing File Management and Organisation

## UNIT 1: CONCEPTS OF DATA PROCESSSING

**CONTENTS**



## 1.0 INTRODUCTION

Data Processing (data processing) is a data collection, storage, retrieval, processing, transformation and transmission. Data is facts, concepts, or instructions of a form of expression, can be manual or automated device for processing in form of digital data, text, graphics or sounds. Data are explained and, after giving a certain sense, they become information. The basic purpose of data processing from a large number, possibly chaotic, difficult to understand and extract the data derived for some specific people who are valuable and meaningful data. Data processing is a system engineering and automatic control of basic aspects. Data processing is applied throughout the social production and social life in various fields. Data processing technology and its application breadth and depth greatly affect the process of development of human society. No one be effective in any life's endeavour without the support of data processing software, data processing software includes: writing process for various programming languages and compilers, data management, file systems and database systems, and various methods of data processing software package. To ensure data security and reliable, a range of data security technologies and management is very essential.

## 2.0 OBJECTIVES

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

- Define the term data processing
- Explain the components of data processing
- List functions of data processing
- Explain at least three types of data

## 3.0 MAIN CONTENT

### 3.1 Definition of Data, Information and Data Processing

You must have performed many daily operations that need your skills in data processing. As you wake up in the morning, you think about number of activities to be scheduled for the day, the amount of money at your disposal to execute the activities and number of people to meet to make you achieve your goals for the day. Relevant data are manipulated to get things done.

#### 3.1.1 Data

Data are any facts, numbers, or text that can be processed by a computer. They are the facts and figures collected, analyzed, and summarized for presentation and interpretation. Data can be obtained from existing sources, by conducting an observational study, or by conducting an experiment. All the data collected in a particular study are referred to as the data set for the study. Today, organizations are accumulating vast and growing amounts of data in different formats and different databases. This includes:

• operational or transactional data such as, sales, cost, inventory, payroll, and accounting

• non-operational data, such as industry sales, forecast data, and macro-economic data

• meta data - data about the data itself, such as logical database design or data dictionary definitions

#### 3.1.2 Information

The patterns, associations, or relationships among all this *data* can provide *information*. For example, analysis of retail point of sale transaction data can yield information on which products are selling and when. The need for converting facts into useful information is not a phenomenon of modern life. Throughout history, and even prehistory, people have found it necessary to sort data into forms that were easier to understand. For example, the ancient Egyptians recorded the ebb and flow of the Nile River and used this information to predict yearly crop yields. Today computers convert data about land and water into recommendations to farmers on crop planting. Mechanical aids to computation were developed and improved upon in Europe, Asia, and America throughout the seventeenth, eighteenth, and nineteenth centuries. Modern computers are marvels

of an electronics technology that continues to produce smaller, cheaper, and more pow- erful components.

### 3.1.3 Data Processing

*Data* are a collection of facts — unorganized but able to be organized into useful information. *Processing* is a series of actions or operations that convert inputs into outputs. When we speak of data processing, the input is data, and the output is useful information. What then is data processing? So, we can define *data processing* as a series of actions or operations that converts data into useful information.

According to Carl (1996) data processing is, broadly, "the collection and manipulation of items of data to produce meaningful information. In this sense it can be considered a subset of *information processing*, "the change (processing) of information in any manner detectable by an observer. The term is often used more specifically in the context of a business or other organization to refer to the class of commercial data processing applications. To Ger, Gustav, Jacqui and Diane (2013), it includes data capture, coding, editing, and imputation; deriving new variables; calculating weights; and finalizing the survey data files

Data processing (numerical and non-numerical) includes the analysis of various, sorting, calculating, editing, processing and handling data. The increasing popularity of computers in the field of computer applications, a small proportion of numerical calculation by computer data processing for information management has become a major application. Such as the side of the draw chart management, warehouse management, accounting management, transportation management, IT management, office automation. Geographical data in a large number of existing data in the natural environment (land, water, climate, biological and other resource data), there are a large number of socio-economic data (population, transport, industry and agriculture, etc.), often require comprehensive data processing. Therefore, the need to establish geographic database, the system to collate and store geographic data to reduce redundancy, the development of data processing software, full use of database technology for data management and processing. Research data also are processed to answer research questions and test hypotheses.

For data processing business website: As the site visits is very large, making a number of professional data analysis, data cleansing often have targeted, that is not related to data, important data such as disposed of. Then, the data related to sub-classification of the segments, you can choose, depending on the analysis of demand pattern analysis techniques, such as path analysis, interest in association rules, clustering and so on.

We use the term *data processing system* to include the resources that are used to accomplish the processing of data. There are four types of resources: people, materials, facilities, and equipment. People provide input to computers, operate them, and use their output. Materials, such as boxes of paper and printer ribbons, are consumed in great quantity. Facilities are required to house the

computer equipment, people and materials.

1.1.3 Types of Data

Think about any collected data that you have experience of; for example, weight, sex, ethnicity, job grade, and consider their different attributes. These variables can be described as categorical or quantitative.

Type of data Level of measurement Examples

Nominal = (no inherent order in categories), eye color, ethnicity, diagnosis

Ordinal - (categories have inherent order). E.g. Job grade, age groups

Categorical - Binary Gender (2 categories – Male / Female)

Discrete - (usually whole numbers) e.g. size of household (ratio)

Quantitative (Interval/Ratio)

Continuous - Can, in theory, take any value in a range, although necessarily recorded to a predetermined degree of precision)

Temperature °C/°F (no absolute zero) (interval)

Height, age (ratio)

Other classification of data according to Bousbia and Belamri (2014) are:

1. Qualitative or quantitative data.
2. Personal, administrative and/or demographic data (age, sex, etc.).
3. Answers to psychological questionnaires for measuring users' satisfaction, motivation, skills, cognitive features, etc.
4. Answers to questions and/or test scores of the academic system.
5. Individual interactions with the educational system: from fine grained actions such as mouse click, to high level ones such as number of attempts, the learner browsing pattern, etc. – Social interaction (chat, sent messages, forum participation, etc.).
6. Visual and facial reactions, etc.

## **3.2** Data processing functions

Data processing may involve various processes, including:

Validation – It ensures that supplied data is clean, correct and useful

Sorting – It arranges items in some sequence and/or in different sets.

Summarization – It reduces detail data to its main points.

Aggregation – It combines multiple pieces of data.

Analysis – It involves collection, organization, analysis, interpretation and presentation of data.

Reporting – It lists detail or summary data or computed information.

Classification – It separates data into various categories.

**IN-TEXT QUESTIONS:**

1. Differentiate between data and information
2. Mention four functions of data processing
3. Enumerate three key terms that are related to data processing.

**Components of Data Processing**

1. Basic data processing operations

Five basic operations are characteristic of all data processing systems: inputting, storing, processing, outputting, and controlling. They are defined as follows.

*Inputting* is the process of entering data, which are collected facts, into a data processing system. *Storing* is saving data or information so that they are available for initial or for additional processing. *Processing* represents performing arithmetic or logical operations on data in order to convert them into useful information. *Outputting* is the process of producing useful information, such as a printed report or visual display. *Controlling* is directing the manner and sequence in which all of the above operations are performed.

2. Data storage hierarchy

It is known that data, once entered, are organized and stored in successively more comprehensive groupings. Generally, these groupings are called a data storage hierarchy. The general groupings of any data storage hierarchy are as follows:

1) *Characters,* which are all written language symbols: letters, numbers, and special symbols. 2) *Data elements,* which are meaningful collections of related characters. Data elements are also

called data items or fields. 3) *Records,* which are collections of related data elements. 4) *Files*, which are collections of related records. A set of related files is called a data base or a data bank. You also need to familiarize yourself with these data processing terms that you will come across often during study of this course.

1. Computer Software: the set of instructions that direct the operations of computers;
2. Computer literacy: a part of a computer, entering data into the device;
3. A program: facts unorganized but able to be organized;
4. Data: the output of a data processing system;
5. Data processing: possessing sufficient knowledge of how computers work tools and what they can do to use them as problem-solving;
6. Data processing f) a series of operations that results in the conversion of data system into useful information;
7. Input: an electronic device performing calculations on numerical data;
8. Output: an electronic device accepting the data processing results from the computer and displaying them;
9. Useful information: a set of related files;
10. Data bank: the resources required to accomplish the processing of data. These resources are personnel, material, facilities and equipment.

**ADVANTAGES OF COMPUTER DATA PROCESSING**

Now that we have discussed functions and components of data processing, let us examine its advantages. Computer-oriented data processing systems or just computer data processing systems are not designed to imitate manual systems. They should combine the capabilities of both humans and computers. Computer data processing systems can be designed to take advantage of four capabilities of computers.

*Accuracy.* Once data have been entered correctly into the computer component of a data processing system, the need for further manipulation by humans is eliminated, and the possibility of error is reduced. Computers, when properly programmed, are also unlikely to make computational errors. Of course, computer systems remain vulnerable to the entry by humans of invalid data.

*Ease of communications.* Data, once entered, can be transmitted wherever needed by communications networks. These may be either earth or satellite-based systems. A travel reservations system is an example of a data communications network. Reservation clerks throughout the world may make an enquiry about transportation or lodgings and receive an almost instant response. Another example is an office communications system that provides executives with access to a reservoir of date, called a corporate data base, from their personal microcomputer work stations.

*Capacity of storage.* Computers are able to store vast amounts of information, to organize it, and to retrieve it in ways that are far beyond the capabilities of humans. The amount of data that

can be stored on devices such as magnetic discs is constantly increasing. All the while, the cost per character of data stored is decreasing.

*Speed.* The speed, at which computer data processing systems can respond, adds to their value. For example, the travel reservations system mentioned above would not be useful if cli- ents had to wait more than a few seconds for a response. The response required might be a fraction of a second. Thus, an important objective in the design of computer data processing systems is to allow computers to do what they do best and to free humans from routine, error- prone tasks. The most cost-effective computer data processing system is the one that does the job effectively and at the least cost. By using computers in a cost-effective manner, we will be better able to respond to the challenges and opportunities of our post-industrial, information-dependent society.

1. Inputting a) saving information for further processing;
2. Character b) the process of producing useful information;
3. Database c) meaningful collections of related characters;
4. Data elements d) the most common input device;
5. Controlling e) the part of the computer that receives and stores data for processing;
6. Outputting f) directing the sequence of the operations performed;
7. Memory g) a written language symbol;
8. Record h) a collection of related data elements
9. Keyboard i) a set of related facts;
10. Storing j) the process of entering collected into a data processing system;

## 4.0 Conclusion

The unit 1 you have studied had exposed you to meaning of data, information and data processing.

## 5.0 Summary

Data are any facts, numbers, or text that can be processed by a computer. Data processing (numerical and non-numerical) includes the analysis of various, sorting, calculating, editing, processing and handling data. Advantages of data processing involve accuracy, easy of communication, capacity of storage and speed.

## 6.0 Tutor-marked Assignment

i. Describe the term data processing

ii, Explain five types of data

iii. Discuss five functions of data processing

iv. Enumerate four advantages of computer data processing.

## 7.0 REFERENCES / FURTHER READINGS


Bousbia, N., & Belamri, I. (2014). Which Contribution Does EDM Provide to Computer-Based Learning Environments? In A. Peña-Ayala (Ed.), *Educational Data Mining: Applications and Trends*. Mexico: Springer International Publishing.

David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, James J. Cochran. (2018). *Statistics for Business and Economics.* (Thirteenth Edition, Revised). United States of America: Cengage Learning.

Ger, S. Gustav, H. Jacqui, J. Diane, K. W. (2013). *Designing and Conducting Business Surveys*. New Jersey: JohnWiley & Sons, Inc.

Starck, J.L., Murtagh, F. & Bijaoui, A (2015). Image Processing and Data Analysis: The Multiscale Approach. Washington. Amazon

# UNIT 2: HISTORICAL PERSPECTIVES AND APPLICATION OF DATA PROCESSING

## CONTENTS



## 1.0 INTRODUCTION

In unit 1, you came across some of the key terms that you need to understand to continue in this course. You have studied that data processing involves inputting, processing and outputting. We shall in this unit intimate ourselves with the historical perspectives of data processing and necessary steps to effective data processing would also be discussed.

## 2.0 OBJECTIVES

- At the end of this unit, you should be able to Discuss the historical perspective in data processing

- ❖ List four application of data processing
- ❖ Explain necessary steps for effective data processing

## 3.0 CONTENTS

## 3.1 Historical Perspective of Data Processing

### 3.1.1 Manual data processing

Although widespread use of the term data processing dates only from the nineteen-fifties[3] data processing functions have been performed manually for millennia. For example bookkeeping involves functions such as posting transactions and producing reports like the balance sheet and the cash flow statement. Completely manual methods were augmented by the application of mechanical or electronic calculators. A person whose job it was to perform calculations manually or using a calculator was called a "computer."

The 1850 United States Census schedule was the first to gather data by individual rather than household. A number of questions could be answered by making a check in the appropriate box on the form. From 1850 through 1880 the Census Bureau employed "a system of tallying, which, by reason of the increasing number of combinations of classifications required, became increasingly complex. Only a limited number of combinations could be recorded in one tally, so it was necessary to handle the schedules 5 or 6 times, for as many independent tallies."[4] "It took over 7 years to publish the results of the 1880 census"[5] using manual processing methods.

### 3.1.2 Automatic data processing

The term automatic data processing was applied to operations performed by means of unit record equipment, such as Herman Hollerith's application of punched card equipment for the 1890 United States Census. "Using Hollerith's punch card equipment, the Census Office was able to complete tabulating most of the 1890 census data in 2 to 3 years, compared with 7 to 8 years for the 1880 census. ... It is also estimated that using Herman Hollerith's system saved some $5 million in processing costs"[5] (in 1890 dollars) even with twice as many questions as in 1880.

### 3.1.3 Electronic data processing

Computerized data processing, or Electronic data processing represents the further evolution, with the computer taking the place of several independent pieces of equipment. The Census Bureau first made limited use of electronic computers for the 1950 United States Census, using a UNIVAC I system, delivered in 1952.

### 3.1.4 Further evolution

Data processing (DP) has also previously been used to refer to the department within an organization responsible for the operation of data processing applications. The term data processing has mostly been subsumed under the newer and somewhat more general term information technology (IT). Data processing has acquired a negative connotation, suggesting use of older technologies. As an example, in 1996 the Data Processing Management Association (DPMA) changed its name to Association of Information Technology Professionals. Nevertheless, the terms are roughly synonymous.

## 3.2 Applications

### 3.2.1 Commercial data processing

Commercial data processing involves a large volume of input data, relatively few computational operations, and a large volume of output. For example, an insurance company needs to keep records on tens or hundreds of thousands of policies, print and mail bills, and receive and post payments.

### 3.2.2 Data analysis

In a science or engineering field, the terms data processing and information systems are considered too broad, and the more specialized term data analysis is typically used. Data analysis makes use of specialized and highly accurate algorithms and statistical calculations that are less often observed in the typical general business environment. One divergence of culture between data processing and data analysis is shown by the numerical representations generally used; In data processing, measurements are typically stored as integers, fixed-point or binary-coded decimal representations of numbers, whereas the majority of measurements in data analysis are stored as floating-point representations of rational numbers. For data analysis, packages like SPSS or SAS, or their free counterparts such as DAP, gretl or PSPP are often used.

## 3.3 Steps to Data Processing

Data is an integral part of all business processes. It is the invisible backbone that supports all the operations and activities within a business. Without access to relevant data, businesses would get completely paralyzed. This is because quality data helps formulate effective business strategies and fruitful business decisions. Therefore, the quality of data should be maintained in good condition in order to facilitate smooth business proceedings. In order to enhance business proceedings, data should be made available in all possible forms in order to increase the accessibility of the same.

Data processing refers to the process of converting data from one format to another. It transforms plain data into valuable information and information into data. Clients can supply data in a variety of forms, be it .xls sheets, audio devices, or plain printed material. Data processing services take the raw data and process it accordingly to produce sensible information. The various

applications of data processing can convert raw data into useful information that can be used further for business processes.

Companies and organizations across the world make use of data processing services in order to facilitate their market research interests. Data consists of facts and figures, based on which important conclusions can be drawn. When companies and organizations have access to useful information, they can utilize it for strategizing powerful business moves that would eventually increase the company revenue and decrease the costs, thus expanding the profit margins. Data processing ensures that the data is presented in a clean and systematic manner and is easy to understand and be used for further purposes.

Here are the 5 steps that are included in data processing:

Editing

There is a big difference between data and useful data. While there are huge volumes of data available on the internet, useful data has to be extracted from the huge volumes of the same. Extracting relevant data is one of the core procedures of data processing. When data has been accumulated from various sources, it is edited in order to discard the inappropriate data and retain relevant data.

Coding

Even after the editing process, the available data is not in any specific order. To make it more sensible and usable for further use, it needs to be aligned into a particular system. The method of coding ensures just that and arranges data in a comprehendible format. The process is also known as netting or bucketing.

Data Entry

After the data has been properly arranged and coded, it is entered into the software that performs the eventual cross tabulation. Data entry professionals do the task efficiently.

**IN-TEXT QUESTIONS**

1. Mention 3 key developments in data processing.
2. What is validation in data processing?
3. Data processing ensures that the data is presented in a clean and systematic manner and is easy to understand and be used for further purposes. True or False

**Validation**

After the cleansing phase, comes the validation process. Data validation refers to the process of thoroughly checking the collected data to ensure optimal quality levels. All the accumulated data is double checked in order to ensure that it contains no inconsistencies and is utterly relevant.

**Tabulation**

This is the final step in data processing. The final product i.e. the data is tabulated and arranged in a systematic format so that it can be further analyzed. All these processes make up the complete data processing activity which ensures the said data is available for access.

## 4.0 CONCLUSION

Data processing had passed through stages of development in which electronic data processing had taken over all forms of manual data processing. Data processing include the steps of data entry, editing, coding, validation and tabulation.

## 5.0 SUMMARY

In this unit you have learnt about the historical evolution of data processing and the steps involved in data processing. The discussion in this unit had also helped you to appreciate how data processing can be carried out in various organisation. Evolvement of data processing comprised manual, automated and electronic operations to put data into meaning form. Data processing is basically applied to data analysis and for commercial purposes. You will come across hard ware components of data processing in unit 3.

### 6.0 TUTOR-MARKED ASSIGNMENT

i. Explain 3 stages of evolution of electronic data processing

ii. Describe 5 steps that are essential for effective data processing

iii List four applications of modern-day data processing.

## 7.0 REFERENCES / FURTHER READINGS


Illingworth, Valerie (2016). A *Dictionary of Computing*. Oxford Paperback Reference (4th ed.). Oxford University Press.

Truesdell, L. E. (1965). *Theory of computer science*.

Bohme, F.; Wyatt, J. P.; Curry, J. P. (1991). *100 Years of Data Processing: The Punch card Century*. United States Bureau of the Census.

Lingworth, V. (1997). *Dictionary of Computing*. Oxford Paperback Reference (4th ed.). Oxford University Press. ISBN 9780192800466.

# UNIT 3: HARDWARE SUPPORTS IN DATA PROCESSING

**CONTENTS**



## 1.0 INTRODUCTION

Now that you have covered the concept of data processing and its historical perspectives in units 1 and 2, it is necessary to discuss the hardware supports in data processing. In this unit we shall look at input, processing and output devices for effective data processing in various organisations.

## 2.0 OBJECTIVES

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

- Identify input, processing and output components of data processing
- Discuss the anatomy of a computer system
- Mention specific examples of output devices
- Differentiate between plotters and printers

## 3.0 CONTENTS

## 3.1. Input, Processing and output Components of Data Processing

Whenever a computer is used it must work its way through three basic stages before any task can be completed. These are input, processing and output. A Computer works through these stages by running a program. A program is a set of step-by-step instructions which tells the computer exactly what to do with the input in order to produce the required output.

**3.1.1** Input

The input stage of computing is concerned with getting the data needed by the program into the computer. Input devices are used to do this. The most commonly used input devices are the mouse and the keyboard.

**3.1.2** Processing

The program contains instructions about what to do with the input. During the processing stage the compute follows these instructions using the data which has just been input. What the computer produces at the end of this stage, the output, will only be as good as the instructions given in the program. In other words if garbage has been put in to the program, garbage is what will come out of the computer. This is known as GIGO, or Garbage In Garbage Out.

**3.1.3** Output

The output stage of computing is concerned with giving out processed data as information in a form that is useful to the user. Output devices are used to do this. The most commonly used output devices are the screen, which is also called a monitor or VDU and the printer.

## 3.2. Architecture of Computer System

This is the 'brain' of the computer. It is where all the searching, sorting, calculating and decision making takes place. The CPU collects all of the raw data from various input devices (such a keyboard or mouse) and converts it into useful information by carrying out software instructions. The result of all that work is then sent to output devices such as monitors and printers.

The CPU is a microprocessor - a silicon chip - composed of tiny electrical switches called 'transistors'. The speed at which the processor carries out its operations is measured in megahertz (MHz) or Gigahertz (GHz). The higher the number of MHz the faster the computer can process information. A common CPU today runs at around 3 GHz or more.

The Intel Pentium processor and the Athlon are examples of a CPU.

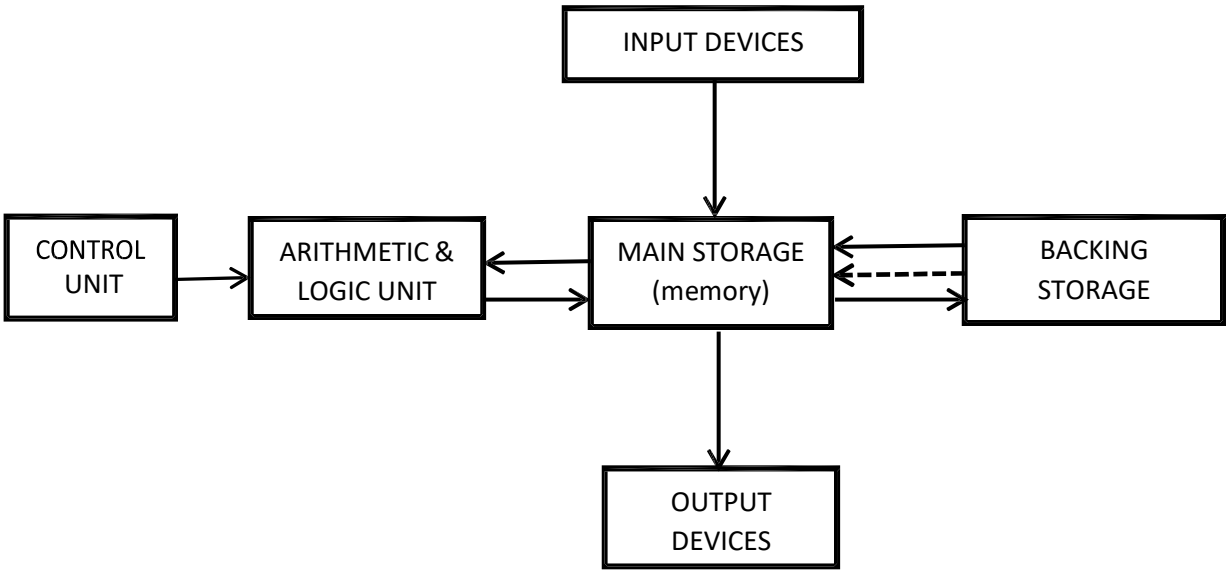



**Figure 1.1 Block diagram of CPU**

### 3.2.1 The Control Unit (CU)

The Control Unit (CU) co-ordinates the work of the whole computer system. It has three main jobs:

1. It controls the hardware attached to the system. The Control Unit monitors the hardware to make sure that the commands given to it by the current program are activated.

2. It controls the input and output of data, so all the signals go to the right place at the right time.

3. It controls the flow of data within the CPU.

### 3.2.2 The Immediate Access Store (IAS)

The Immediate Access Store (IAS) holds the data and programs needed at that instant by the Control Unit. The CPU reads data and programs kept on the backing storage and store them temporarily in the IAS's memory. The CPU needs to do this because Backing Store is much too slow to be able to run data and programs from directly. For example, lets pretend that a modern CPU was slowed down to carry out one instruction in 1 second, then the hard disk (i.e. Backing Store) would take 3 months to supply the data it needs!

So the trick is to call in enough of the data and programs into fast Immediate Access Store memory so as to keep the CPU busy.

### 3.2.3 ALU stands for Arithmetic and Logic Unit.

It is where the computer processes data by either manipulating it or acting upon it. It has two parts:

1. Arithmetic part - does exactly what you think it should - it does the calculations on data such as 3 + 2.

2. Logic part - This section deals with carrying out logic and comparison operations on data. For example working out if one data value is bigger than another data value.

### 3.2.4 Input Devices

Due to a constant research in the computer hardware we have a large number of input devices recall that before data can be processed by the computer they must be translated into machine readable form and entered into the computer by an input device. Here we will introduce a variety of input devices.

#### 3.2.4.1 Keyboard

The keyboard is the most widely used input device and is used to enter data or commands to the computer. It has a set of alphabet keys, a set of digit keys, and various function keys and is divided into four main areas:

• Function keys across the top

• Letter keys in the main section

• A numeric keypad on the right

• Cursor movement and editing keys between the main section and the numeric keypad.

The layout of the letters on a keyboard is standard across many countries and is called a QWERTY keyboard. The name comes from the first six keys on the top row of the alphabetic characters. Some keyboards come with added keys for using the Internet and others have an integrated wrist support. Ergonomic keyboards have been developed to reduce the risk of repetitive strain injury to workers who use keyboards for long periods of time.

The computer's processor scans the keyboard hundreds of times per second to see if a key has been pressed. When a key is pressed, a digital code is sent to the Central Processing Unit (CPU). This digital code is translated into ASCII code (American Standard Code of Information Interchange).

For example, pressing the 'A' key produces the binary code 01100001 representing the lower case letter 'a'. Holding down the shift key at the same time produces the binary code 01000001 representing the upper case letter 'A'.

Advantages:

• Most computers have this device attached to it

• It is a reliable method for data input of text and numbers

• A skilled typist can enter data very quickly.

• Specialist keyboards are available

Disadvantages:

• It is very easy to make mistakes when typing data in

• It can be very time consuming to enter data using a keyboard, especially if you are not a skilled typist.

• It is very difficult to enter some data, for example, details of diagrams and pictures.

• It is very slow to access menus and not flexible when you want to move objects around the screen

• Difficult for people unable to use keyboards through paralysis or muscular disorder.

**3.2.4.2 Mouse**

A mouse is the most common pointing device that you will come across. It enables you to control the movement and position of the on-screen cursor by moving it around on the desk.

Buttons on the mouse let you select options from menus and drag objects around the screen. Pressing a mouse button produces a 'mouse click'. You might have heard the expressions 'double click', 'click and drag' and 'drag and drop'.

Most mice use a small ball located underneath them to calculate the direction that you are moving the mouse in. The movement of the ball causes two rollers to rotate inside the mouse; one records the movement in a north-south direction and the other records the east-west movement. The mouse monitors how far the ball turns and in what direction and sends this information to the computer to move the pointer.

Advantages:

• Ideal for use with desktop computers.

• Usually supplied with a computer so no additional cost.

• All computer users tend to be familiar with using them.

Disadvantages:

 They need a flat space close to the computer.

• The mouse cannot easily be used with laptop, notebook or palmtop computers. (These need a tracker ball or a touch sensitive pad called a touch pad).

**3.2.4.3 Trackball**

A tracker ball is like an upside-down mouse with the ball on top. Turning the ball with your hand moves the pointer on the screen. It has buttons like a standard mouse, but requires very little space

to operate and is often used in conjunction with computer aided design. You will often find a small tracker ball built into laptop computers in place of the conventional mouse.

Advantages:

• Ideal for use where flat space close to the computer is limited.

• Can be useful with laptops as they can be built into the computer keyboard or clipped on.

Disadvantages:

• Not supplied as standard so an additional cost and users have to learn how to use them

**3.2.4.4 Joystick**

A Joystick is similar to a tracker ball in operation except you have a stick which is moved rather than a rolling ball. Joysticks are used to play computer games. You can

move a standard joystick in any one of eight directions. The joystick tells the computer in which direction it is being pulled and the computer uses this information to

(for example) move a r acing car on screen. A joystick may also have several buttons which can be pressed to trigger actions such as firing a missile.

Advantages:

• There is an immediate feel of direction due to the movement of the stick

Disadvantages:

• Some people find the joystick difficult to control rather than other point and click devices. This is probably because more arm and wrist movement is required to control the pointer than with a mouse or tracker ball.

• Joysticks are not particularly strong and can break easily when used with games software.

**3.2.4.5 Touch Screen**

These screens do a similar job to concept keyboards. A grid of light beams or fine wires crisscross the computer screen. When you touch the screen with your finger, the rays are blocked and the computer 'senses' where you have pressed. Touch screens can be used to choose options which are displayed on the screen.

Touch screens are easy to use and are often found as input devices in public places such as museums, building societies (ATMs), airports or travel agents. However, they are not commonly used elsewhere since they are not very accurate, tiring to use for a long period and are more expensive than alternatives such as a mouse.

Advantages:

• Easy to use

• Software can alter the screen while it is running, making it more flexible that a concept keyboard with a permanent overlay

• No extra peripherals are needed apart from the touch screen monitor itself.

• No experience or competence with computer systems are needed to be able to use it.

Disadvantages:

• Not suitable for inputting large amounts of data

• Not very accurate, selecting detailed objects can be difficult with fingers

• Tiring to use for a long period of time

• More expensive than alternatives such as a mouse.

• Touch screens are not robust and can soon become faulty.

**3.2.4.6 Digital Camera**

A digital camera looks very similar to a traditional camera. However, unlike photographic cameras, digital cameras do not use film. Inside a digital camera is an array of light sensors. When a picture is taken, the different colors that make up the picture are converted into digital signals (binary) by sensors placed behind the lens.

Most digital cameras let you view the image as soon as you have taken the picture and, if you don't like what you see, it can be deleted. The image can then be stored in the camera's RAM or on a floppy disk. Later, the pictures can be transferred onto a computer for editing using photo imaging software. The amount of memory taken up by each picture depends on its resolution.

The resolution is determined by the number of dots which make up the picture: the greater the number of dots which make up the picture, the clearer the image. However, higher resolution pictures take up more memory (and are more expensive).

Resolution range from about 3 million (or Mega) pixels up to 12 Mega pixels Digital cameras are extremely useful for tasks such as producing newsletters. There is often a digital camera built into mobile phones that operates in exactly the same way as a standard one.

Advantages:

• No film is needed and there are no film developing costs

• Unwanted images can be deleted straight away

• You can edit, enlarge or enhance the images

• Images can be incorporated easily into documents, sent by e-mail or added to a website.

Disadvantages:

• Digital cameras are generally more expensive than ordinary cameras.

• Images often have to be compressed to avoid using up too much expensive memory

• When they are full, the images must be downloaded to a computer or deleted before any more can be taken.

**3.2.4.7 Scanner**

A scanner is another way in which we can capture still images or text to be stored and used on a computer. Images are stored as 'pixels'.

A scanner works by shining a beam of light on to the surface of the object you are scanning. This light is reflected back on to a sensor that detects the color of the light. The reflected light is then digitized to build up a digital image. Scanner software usually allows you to choose between a high resolution (very high quality images taking up a lot of memory) and lower resolutions.

Special software can also be used to convert images of text into actual text data which can be edited by a word processor. This software is called an "Optical Character Reader" or OCR.

There are two types of scanner:

• Flatbed Scanner

• Handheld Scanner

The most popular type of scanner is the flatbed. It works in a similar way to a photocopier. Flatbed scanners can scan larger images and are more accurate than handheld scanners.

Handheld scanners are usually only a few inches wide and are rolled across the document to be scanned. They perform the same job but the amount of information that can be scanned is limited by the width of the scanner and the images produced are not of the same quality as those produced by flatbed scanners.

Advantages:

• Flat-bed scanners are very accurate and can produce images with a far higher resolution than a digital camera

• Any image can be converted from paper into digital format and later enhanced and used in other computer documents.

Disadvantages:

• Images can take up a lot of memory space.

• The quality of the final image depends greatly upon the quality of the original document.

#### 3.2.4.8 Graphics Tablets

Graphics tablets are often used by graphics designers and illustrators. Using a graphics tablet a designer can produce much more accurate drawings on the screen than they could with a mouse or other pointing device. A graphics tablet consists of a flat pad (the tablet) on which you draw with a special pen. As you draw on the pad the image is created on the screen. By using a graphics tablet a designer can produce very accurate on-screen drawings.

Drawings created using a graphics tablet can be accurate to within hundredths of an inch. The 'stylus' or pen that you use may have buttons on it that act like a set of mouse buttons. Sometimes, instead of a stylus a highly accurate mouse-like device called a puck is used to draw on the tablet.

Advantages:

• In a design environment where it is more natural to draw diagrams with pencil and paper, it is an effective method of inputting the data into the computer.

Disadvantages:

• Not as good as a mouse for clicking on menu items.

### 3.2.5 Output Devices

Once data has been input into a computer and processed, it is of little use unless it can be retrieved quickly and easily from the system. To allow this, the computer must be connected to an output device. The most common output devices are computer monitors and printers.

However, output can also be to a modem, a plotter, speakers, a computer disk, another computer or even a robot.

#### 3.2.5.1 Monitor

A Monitor (or "screen") is the most common form of output from a computer. It displays information in a similar way to that shown on a television screen. On a typical computer the monitor may measure 17 inches (43 cm) across its display area. Larger monitors make working at a computer easier on the eyes. Of course, the larger the screen, the higher its cost! Typical larger sizes are 19-inch, 20 inch and 21 inches.

Part of the quality of the output on a monitor depends on what resolution it is capable of displaying. Other factors include how much contrast it has, its viewing angle and how fast does it refresh the screen. For example, a good computer game needs a fast screen refresh so you can see all the

action. The picture on a monitor is made up of thousands of tiny colored dots called

pixels. The quality and detail of the picture on a monitor depends on the number of pixels that it can display. The denser the pixels the greater the clarity of the screen image.

A PC monitor contains a matrix of dots of Red, Green and Blue known as RGB. These can be blended to display millions of colors.

This is one RGB pixel of light

R + B = M (magenta)

B + G = C (cyan)

G + R = Y (yellow)

R + G + B = W (white)

The two most common types of monitor are a cathode-ray tube (CRT) monitor and a liquid crystal display (LCD).

Liquid Crystal Display (or "TFT" Display)

This is smaller and lighter than the CRT (see below), which makes them ideal for use with portable laptops, PDAs and Palmtops. Even desktop computers are using them now that their price has become comparable to CRT monitors.

Liquid Crystal is the material used to create each pixel on the screen. The material has a special property - it can 'polarize' light depending on the electrical charge across it. Charge it one way and all the light passing through it is set to "vertical" polarity, charge it another way and the light polarity is set to "horizontal". This feature allows the pixels to be created. Each tiny cell of liquid crystal is a pixel.

TFT (or Thin Film Transistor) is the device within each pixel that sets the charge. And so sometimes they are called "Liquid Crystal Display" referring to the material they use or they are called "TFT displays" referring to the tiny transistors that make them work.

LCDs use much less power than a normal monitor.

Cathode Ray Tube

The CRT works in the same way as a television - it contains an electron gun at the back of the glass tube. This fires electrons at groups of phosphor dots which coat the inside of the screen. When the electrons strike the phosphor dots they glow to give the colors.

Advantages of monitors

• Relatively cheap

• Reliable

• Can display text and graphics in a wide range of colours

• As each task is processed, the results can be displayed immediately on the screen

• Output can be scrolled backwards and forwards easily.

• Quiet

• Do not waste paper

Disadvantages of monitors:

• No permanent copy to keep - the results will disappear when the

computer is switched off.

• Unsuitable for users with visual problems.

• Only a limited amount of information can be displayed at any one time

• Screens are made of glass and can be very fragile.

### 3.2.5.2 Printers

Printers are output devices. They are dedicated to creating paper copies from the computer. Printers can produce text and images on paper. Paper can be either separate sheets such as A4 A5 A3 etc. or they may be able to print on continuous

(fanfold) paper that feed through the machine. A ream of A4 paper continuous paper with holes on the edges, used by dot matrix printers. After you print on fanfold paper, you have to separate the pages and tear off the edge strips. Very specialist printers can also print on plastic or even textiles such as T shirts.

Some printers are dedicated to only producing black and white output. Their advantage is that they are often faster than a color printer because effectively there is only one color to print (Black). Color Printers are dedicated to creating text and images in full

color. Some types can even produce photographs when special paper is used.

There are three main types of printer that you need to know about. You will be expected to understand the main differences i.e. purchase costs, running costs, quality and speed

The three types are Laser, Dot Matrix and Inkjet.

### 3.2.5.3 Plotter

These are output devices that can produce high quality line diagrams on paper. They are often used by engineering, architects and scientific organizations to draw plans, diagrams of machines and printed circuit boards.

A plotter differs from a printer in that it draws images using a pen that can be lowered, raised and moved across the page to form continuous lines. The electronically controlled pen is moved by two computer-controlled motors. The pen is lifted on and off the page by switching an electromagnet on and off.

The paper is handled in different ways depending on the type of plotter.

Flatbed plotters hold the paper still while the pens move.

Drum plotters roll the paper over a cylinder Pinch-roller plotter with a mixture of the two.

Advantages:

• Drawings are of the same quality as if an expert drew them

• Larger sizes of paper can be used than would be found on most printers

Disadvantages:

• Plotters are slower than printers, drawing each line separately.

• They are often more expensive to buy than printers

• Although drawings are completed to the highest quality they are not suitable for text (although text can be produced)

• There is a limit to the amount of detail these plotters can produce, although there are plotters which are "pen-less" the set are used for high-density drawings as may be used for printed circuit board layout.

• In recent years, cheaper printers that can handle A3 and A2 sized paper have resulted in a decline in the need for smaller plotters.

**IN-TEXT QUESTION S**

Reflect on what you have studied by answering the following questions

• What do you mean by information? How it is different from data? Explain.
• Explain the process of input – processing - output with the help of suitable examples.
• Explain the architecture of a Computer System.

## 4.0 CONCLUSION

This unit focused on hardware devices for data processing. We need relevant input output devices to enter data and to show the results of every processes that is carried out by the computer. Your knowledge of computer hardware had given you basic skills to use relevant devices to input, process and obtain information necessary foe decision-making in your educational endeavor and business ventures.

### 5.0 SUMMARY

Data processing is any computer process that converts data into information. Data are any facts, numbers, or text that can be processed by a computer. The patterns, associations, or relationships among all this *data* can provide *information*. The CPU is a microprocessor - a silicon chip - composed of tiny electrical switches called 'transistors'. The keyboard is the most widely used input device and is used to enter data or commands to the computer. A Joystick is similar to a tracker ball in operation except you have a stick which is moved rather than a rolling ball. Graphics tablets are often used by graphics designers and illustrators. The most common output devices are computer monitors and printers. Meta data is data about the data itself, such as logical database design or data dictionary definitions. Resolution of a digital camera range from about 3 million (or Mega) pixels up to 12 Mega pixels.

## 6.0 TUTOR-MARKED ASSIGNMENT

1. Explain what is meant by the term input device? Give three examples of input devices. Also give possible advantages and disadvantage of the same.

2. Explain what is meant by the term output device? Give three examples of output devices. Also give possible advantages and disadvantage of the same.

3. What are different types of printers? How a plotter is different from a printer?

### 7.0 REFERENCES / FURTHER READINGS

Clifton, H.D. (2017). *Business Data Systems: A Practical Guide to Systems Analysis and Data Processing*. London: Prentice Hill

## UNIT 4: SOFTWARE ELEMENTS IN DATA PROCESSING

**CONTENTS**



### 1.0 INTRODUCTION

You have learnt from the previous unit that hardware are equipment used for data processing. They operate in form of input, processing and output units. The equipment are useless unless some other materials are installed into them to make them work. In this unit therefore, you will learn about software that drive the hardware to function.

### 2.0 OBJECTIVES

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

- Define software
- Differentiate between system software and application software
- List the functions of operating system
- Enumerate the considerations to be made in selecting application software

## 3.0 MAIN CONTENTS

### 3.1 Definition and Types of Software

Software is suite programs that make the hardware components of computers to function. It is a collection of programs. Programs are computer instructions which control the works of

computers. These are computer languages which guides and directs how computers are put into use. Software are categorized into two broad ways: System software and Application software. Let us look at each of them one by one.

### 3.1.1 Operating system

Operating system are computer-oriented software which controls the internal operation of computer system. An operating system consists of suites of programs , one of which the master, kernel or executive program, remains resident in the main store of the computer. This program controls the other operating system programs in the suite and between them they control the application programs. Example of current operating system that are commonly used today is Windows coming up in various versions.

The following are the functions of operating system:

Priority assignment: Operating system manipulates jobs awaiting execution are scheduled according to either a predetermined or dynamic assignment plan.

Control of multi-programming in which programs are executed in order to achieve their purposes.

Spooling: The control of input/output peripheral in order to achieve their utilization.

Communication: It controls data transmission between terminals and computer, and computer-to-computer.

Dynamic allocation of main and backing storage, including virtual storage

Database: It also controls database management system.

Software control: Software such as assemblers, compilers, utility software and subroutines are controlled so that they available immediately when required.

Application package control especially with minicomputers.

Operations log: This involves maintenance of details of all jobs carried out by the computer

## 3.2 Application Software

### 3.2.1 Description of Application Software

Application software comprises the programs that are written specifically to achieve results appertaining to the company's activities. In other words application software is user-oriented as opposed to systems software which computer-oriented. Application software comes from two sources, i.e. the company's own DP staff or from external agencies. In the early years of DP, purchased software was often found to be too restrictive and badly documented so that apart

from payroll was not accepted. Existing application packages are available for use. A business application package is a complete suite of programs together with associated documentation. It covers a business routine, is supplied by a computer manufacturer or software house, on lease or purchase. A package is normally intended to meet the needs of a number of different user companies. It also contains a number of options; these are selected by user by the insertion of parameters before use.

### 3.2.2 Advantages of Application Packages

The following benefits should be accrued from the adoption of an application package:

Implementation of an application is quicker and possibly cheaper.

System designs, programming and system testing are maximised.

System documentation is provided with the package.

Portability from the existing computer system to any new computer adopted by the user.

Efficiency in terms of speed, storage requirements and accuracy.

## 3.3 Considerations Regarding Application Packages.

Before any organization considers the use of application package, it is necessary to give consideration to:

Definition of Requirements

The user company cannot abandon the study of company objectives, systems investigation and consequent definition of DP requirements. In these respects the approach is the same as when designing a DP systems for in-house programming.

Study of Range

A range of packages should be examined in depth before a choice is made, and existing users of the packages should be consulted for their practical experiences and opinions. It should be remembered that the more commonly used packages result in pressure on the

suppliers to keep them up-to-date.

Interfacing

Consideration should be given to how a package interface with the user's own routines, both existing and future especially in the area of database.

Amendments

At least one member of the user company staff must be completely conversant with the detailed operations and capabilities of the package adopted. The package supplier must be in a position to supply all the necessary on-going amendments, especially those required through legislation such as for taxation.

Performance

How efficient is the package in terms of its average and maximum run times on the computer? What resources does it demand such as peripherals and of main and backing storage?

Contract Terms

The terms of contract should embrace factors such as the terms of payment, suppliers' assistance with implementation, extent of documentation and future maintenance.

## 3.4 Types of Application Software

### 3.4.1 Spreadsheet Software

A spreadsheet also known as a worksheet, is a multipurpose method that is usable for variety of planning, modeling and forecasting purposes, e.g. budgeting, sales analysis, payroll and break-even analysis. The principle of a spreadsheet is that it stimulates a large matrix of cells within each of which a data item or formula can be held. A display of cells, i.e. a window, is scrolled up or down and right or left so that any cell may be inspected at will. Each are located using the row and column identifier. All vertical cells are referred to as columns while the horizontal cells are rows. Among the better-known spreadsheet programs are Microsoft Excel, Multiplan, Supercalc, Lotus Works and Visicalc.

### 3.4.2 Word processing Software

Word processing is an enhanced method of typing which converts text into form the typewritten documents but also stores it in computer storing devices like hard disks, floppy drives, flash drives, CD-ROM and tapes (Opateye, 1999). It involves the composition, recording, transcription, editing and communication of text. A feature of Word processing is its ability to interface with other text that automatically incorporates the relevant names, descriptions and numerical data. Word processing used in offices can be categorized into the following groups:

One-off texts, such as individualized letters and memoranda

Replicated texts, such as standard letters and certain legal documents, perhaps with some degree of individuality.

Updatable texts, such as reports and manufacturers' manuals that may need amending at regular and frequent intervals.

Examples of Word processing packages are Windows Microsoft Word, Windows Word-perfect and Page maker.

### 3.4.2 Database Management Software (DBMS)

A database is a collection of data supporting the operations of the organization. More specifically, a database entails creating and maintaining data in computer storage in such a way that it is usable for many purposes. In order to have an efficient database there are certain characteristics that must be met, i.e. it must:

Be substantially non-redundant

Be program independent

Be usable by all the programs

Include all the necessary structural interrelationships of data

Have common approach to the retrieval, insertion and amendment of data

Database management system therefore could be defined as a system that organizes the storage of data in such a way to facilitate its retrieval for many different applications. The characteristics of logical data are decided by the user's need for flexibility in processing data. In many business situations there is an unlimited range of demands for data; new problems and consequent new arrangements of data occur frequently. Examples of facilities of DBMS are

Screen formatting for ease of data entry

Record and file locking to make multi-user systems secure

Sorting records into any sequence

Creation of an audit trail

Logging of transactions

Control of Passwords

Acceptance of written programs for enhancing DBMS

Acceptance of high-level languages as a means of making changes to the contents of the database

Validation of data

Dynamic creation and maintenance of data dictionary

DBMS application packages available for use today are Oracles, FoxPro, Microsoft Access etc.

3.4.4 Accounting Software

This involve a wide range of selection of software that is available for business applications such as payroll, sales, purchases and nominal ledgers, stock control, financial modeling, survey and so on.

3.4.5 Statistical Analyses Software

Statistical analysis software is an application package that helps in processing research-based data in order to answer research questions and test hypotheses. Such software helps in entering and processing data collect during the course of study. Data analyses packages include Statistical Package for Social Scientist (SPSS), Statistical Analysis for Scientists (SAS) among others.

**IN-TEXT QUESTIONS**

Reflect on what you have studied in this unit by answering the following questions

- What do you understand by "Software"?
- Application software are programs written to control all operations of computer. Yes / No
- List three requirements to be considered when you want to acquire an application package.

## 4.0 CONCLUSION

Software play important roles in the operations of computer. None of the hardware components of the computer could efficiently utilized without software that enable them to work. Software could be system in nature which are machine language or operating system that controls all operations and application packages that are user-oriented b ased on the demands of activities of the user. Certain considerations should be put in place before purchasing or developing application software.

## 5.0 SUMMARY

In unit 4, we have looked at software as suite programs that make the hardware components of computers to function. It is a collection of programs. Programs are computer instructions which control the works of computers. Operating system are computer-oriented software which controls the internal operation of computer system. application software is user- oriented as opposed to systems software which computer-oriented.

## 6.0 TUTOR-MARKED ASSIGNMENT

i. Mention five functions of Operating system.

ii. Enumerate five advantages of application software

iii. Differentiate between Spreadsheet and Word processing

iv. Distinguish between database and database management system.

## 7.0 REFERENCES / FURTHER READINGS

Opateye, J.A. (2015). *Elementary Computer Studies for Schools.* Lagos. Arise Publications.

Carl, F. (2014). Data Processing and Information Technology. London. Thomson

## UNIT 5: DATA PROCESSING FILE MANAGEMENT AND ORGANISATION



## 1.0 INTRODUCTION

A computer file is a piece of arbitrary information, or resource for storing information, that is available to a computer program and is usually based on some kind of durable storage. A file is *durable* in the sense that it remains available for programs to use after the current program has finished. Computer files can be considered as the modern counterpart of the files of printed documents that traditionally existed in offices and libraries.

## 2.0 OBJECTIVES

At the conclusion of this unit you should be able to

- Describe content of a computer file
- List 4 operations that can be performed on the file
- Differentiate between relative and indexed files
- Discuss different data capturing techniques

# 3. CONTENTS

## 3.1 Concept and Contents of Files

As far as the operating system is concerned, a file is in most cases just a sequence of binary digits. At a higher level, where the content of the file is being considered, these binary digits may represent integer values or text characters, It is up to the program using the file to understand the meaning and internal layout of information in the file and present it to a user as a document, image, song, or program.

At any instant in time, a file has might have a size, normally expressed in bytes, that indicates how much storage is associated with the file. Information in a computer file can consist of smaller packets of information (often called *records* or *lines*) that are individually different but share some trait in common. For example, a payroll file might contain information concerning all the employees in a company and their payroll details; each record in the payroll file concerns just one employee, and all the records have the common trait of being related to payroll—this is very similar to placing all payroll information into a specific filing cabinet in an office that does not have a computer. A text file may contain lines of text, corresponding to printed lines on a piece of paper.

The way information is grouped into a file is entirely up to the person designing the file. This has led to a plethora of more or less standardized file structures for all imaginable purposes, from the simplest to the most complex. Most computer files are used by computer programs. These programs create, modify and delete files for their own use on an as-needed basis. The programmers who create the programs decide what files are needed, how they are to be used and (often) their names.

In some cases, computer programs manipulate files that are made visible to the computer user. For example, in a word-processing program, the user manipulates document files that she names herself. The content of the document file is arranged in a way that the word-processing program understands, but the user chooses the name and location of the file, and she provides the bulk of the information (such as words and text) that will be stored in the file.

Files on a computer can be created, moved, modified, grown, shrunk and deleted. In most cases, computer programs that are executed on the computer handle these operations, but the user of a computer can also manipulate files if necessary. For instance, Microsoft Word files are normally created and modified by the Microsoft Word program in response to user commands, but the user can also move, rename, or delete these files directly by using a *file manager program* such as Windows Explorer (on Windows computers).

## 3.2 File Operations and Organisation

### 3.2.1 Operations on the file

*Opening* a file to use its contents

*Reading* or *updating* the contents

*Committing* updated contents to durable storage

*Closing* the file, thereby losing access until it is opened again

### 3.2.2 File Organization

#### 3.2.2.1 Sequential file

Access to records in a Sequential file is serial. To reach a particular record, all the preceding records must be read. As we observed when the topic was introduced earlier in the course, the organization of an unordered Sequential file means it is only practical to read records from the file and add records to the end of the file (OPEN.EXTEND).

It is not practical to delete or update records. While it is possible to delete, update and insert records in an ordered Sequential file, these operations have some drawbacks.

Problems accessing ordered Sequential files

Records in an ordered Sequential file are arranged, in order, on some key field or fields. When we want to insert, delete or amend a record we must preserve the ordering. The only way to do this is to create a new file. In the case of an insertion or update, the new file will contain the inserted or updated record. In the case of a deletion, the deleted record will be missing from the new file.

The main drawback to inserting, deleting or amending records in an ordered Sequential file is that the entire file must be read and then the records written to a new file. Since disk access is one of the slowest things we can do in computing this is very wasteful of computer time when only a few records are involved.

For instance, if 10 records are to be inserted into a 10,000 record file, then 10,000 records will have to be read from the old file and 10,010 written to the new file. The average time to insert a new record will thus be very great.

Inserting records in an ordered Sequential file

To insert a record in an ordered Sequential file:

1. All the records with a key value less than the record to be inserted must be read and then written to the new file.

2. Then the record to be inserted must be written to the new file.

3. Finally, the remaining records must be written to the new file. Deleting records from an ordered Sequential file

To delete a record in an ordered Sequential file:

1. All the records with a key value less than the record to be deleted must be written to the

new file.

2. When the record to be deleted is encountered it is not written to the new file.

3. Finally, all the remaining records must be written to the new file.

Amending records in an ordered Sequential file

To amend a record in an ordered Sequential file:

1. All the records with a key value less than the record to be amended must be read and then written to the new file.

2. Then the record to be amended must be read the amendments applied to it and the amended record must then be written to the new file.

3. Finally, all the remaining records must be written to the new file.

### 3.2.2.2 Relative File

As we have already noted, the problem with Sequential files is that access to the records is serial. To reach a particular record, all the proceeding records must be read. Direct access files allow direct access to a particular record in the file using a key and this greatly facilitates the operations of reading, deleting, updating and inserting records. COBOL supports two kinds of direct access file organizations -Relative and Indexed.

Organization of Relative files

Records in relative files are organized on ascending Relative Record Number.

A Relative file may be visualized as a one dimension table stored on disk, where the Relative Record Number is the index into the table. Relative files support sequential access by allowing the active records to be read one after another. Relative files support only one key. The key must be numeric and must take a value between 1 and the current highest Relative Record Number. Enough room is allocated to the file to contain records with Relative Record Numbers between 1 and the highest record number. For instance, if the highest relative record number used is 10,000 then room for 10,000 records is allocated to the file. In this example, enough room has been allocated on disk for 328 records. But although there is room for 328 records in the current allocation, not all the record locations contain records. The record areas labeled "free", have not yet had record values written to them.

Accessing records in a Relative file

To access a record in a Relative file a Relative Record Number must be provided. Supplying this number allows the record to be accessed directly because the system can use the start position of

the file on disk, the size of the record, and the Relative Record Number to calculate the position of the record. Because the file management system only has to make a few calculations to find the record position the Relative file organization is the fastest of the two direct access file organizations available in COBOL. It is also the most storage efficient.

3.2.2.3 Indexed Files

While the usefulness of a Relative file is constrained by its restrictive key, Indexed files suffer from no such limitation. Indexed files may have up to 255 keys, the keys can be alphanumeric and only the primary key must be unique. In addition, it is possible to read an Indexed file sequentially on any of its keys.

Organization of Indexed files

An Indexed file may have multiple keys. The key upon which the data records are ordered is called the primary key. The other keys are called alternate keys. Records in the Indexed file are sequenced on ascending primary key. Over the actual data records, the file system builds an index. When direct access is required, the file system uses this index to find, read, insert, update or delete, the required record. For each of the alternate keys specified in an Indexed file, an alternate index is built. However, the lowest level of an alternate index does not contain actual data records. Instead, this level made up of base records which contain only the alternate key value and a pointer to where the actual record is. These base records are organized in ascending alternate key order.

As well as allowing direct access to records on the primary key or any of the 254 alternate keys, indexed files may also be processed sequentially. When processed sequentially, the records may be read in ascending order on the primary key or on any of the alternate keys. Since the data records are in held in ascending primary key sequence it is easy to see how the file may be accessed sequentially on the primary key. It is not quite so obvious how sequential on the alternate keys is achieved. This is covered in the unit on Indexed files.

## 3.3 Organizing files and folders

Files and folders arranged in a hierarchy In modern computer systems, files are typically accessed using names. In some operating systems, the name is associated with the file itself. In others, the file is anonymous, and is pointed to by links that have names. In the latter case, a user can identify the name of the link with the file itself, but this is a false analogue, especially where there exists more than one link to the same file.

Files (or links to files) can be located in *directories*. However, more generally, a directory can contain either a list of files, or a list of links to files. Within this definition, it is of paramount importance that the term "file" includes directories. This permits the existence of directory hierarchies. A name that refers to a file within a directory must be unique. In other words, there must be no identical names in a directory. However, in some operating systems, a name may include a specification of type that means a directory can contain an identical name to more than

one type of object such as a directory and a file.

In environments in which a *file* is named, a file's name and the path to the file's directory must uniquely identify it among all other files in the computer system—no two files can have the same name and path. Where a file is anonymous, named references to it will exist within a namespace. In most cases, any name within the namespace will refer to exactly zero or one file. However, any file may be represented within any namespace by zero, one or more names. Any string of characters may or may not be a well-formed name for a file or a link depending upon the context of application. Whether or not a name is well formed depends on the type of computer system being used. Early computers permitted only a few letters or digits in the name of a file, but modern computers allow long names (some up to 255) containing almost any combination of unicode letters or unicode digits, making it easier to understand the purpose of a file at a glance. Some computer systems allow file names to contain spaces; others do not. Such characters such as / or \ are forbidden. Case-sensitivity of file names is determined by the file system.

Most computers organize files into hierarchies using *folders*, *directories*, or *catalogs.* Each folder can contain an arbitrary number of files, and it can also contain other folders. These other folders are referred to as *subfolders*. Subfolders can contain still more files and folders and so on, thus building a tree-like structure in which one "master folder" (or "root folder" — the name varies from one operating system to another) can contain any number of levels of other folders and files. Folders can be named just as files can (except for the root folder, which often does not have a name). The use of folders makes it easier to organize files in a logical way.

### 3.3.1 Protecting files

Many modern computer systems provide methods for protecting files against accidental and deliberate damage. Computers that allow for multiple users implement file permissions to control who may or may not modify, delete, or create files and folders. A given user may be granted only permission to modify a file or folder, but not to delete it; or a user may be given permission to create files or folders, but not to delete them. Permissions may also be used to allow only certain users to see the contents of a file or folder. Permissions protect against unauthorized tampering or destruction of information in files, and keep private information confidential by preventing unauthorized users from seeing certain files. Another protection mechanism implemented in many computers is a *read-only flag.* When this flag is turned on for a file (which can be accomplished by a computer program or by a human user), the file can be examined, but it cannot be modified. This flag is useful for critical information that must not be modified or erased, such as special files that are used only by internal parts of the computer system. Some systems also include a *hidden flag* to make certain files invisible; this flag is used by the computer system to hide essential system files that users must never modify

### 3.3.2 Storing files

In physical terms, most computer files are stored on *hard disks*—spinning magnetic disks inside a computer that can record information indefinitely. Hard disks allow almost instant access to

computer files. On large computers, some computer files may be stored on magnetic tape. Files can also be stored on other media in some cases, such as writeable *compact discs, Zip drives,* etc.

### 3.3.3 Backing up files

When computer files contain information that is extremely important, a *backup* process is used to protect against disasters that might destroy the files. Backing up files simply means making copies of the files in a separate location so that they can be restored if something happens to the computer, or if they are deleted accidentally. There are many ways to back up files. Most computer systems provide utility programs to assist in the back-up process, which can become very time consuming if there are many files to safeguard. Files are often copied to removable media such as writeable CDs or cartridge tapes. Copying files to another hard disk in the same computer protects against failure of one disk, but if it is necessary to protect against failure or destruction of the entire computer, then copies of the files must be made on other media that can be taken away from the computer and stored in a safe, distant location.

### 3.3.4 File Terminology

There are a few terms that you need to understand when learning about file system. These will be explained over the next couple of pages. File can store data or information in various formats. Suppose in a file data is stored in the tables just like the one below:

## 3.4 Records

As you saw previously, each table stores can hold a great deal of data. Each table contains a lot of records. A record is all of the data or information about one person or one thing.

## 3.5 Fields

Each table contains a lot of records. A record is made up of lots of individual pieces of information. Look at Wonder Woman's record; it stores her first name, last name, address, city and age. Each of these individual pieces of information in a record are called a 'field'. A 'field' is one piece of data or information about a person or thing.

What fields do you think would be stored in your student record on the school database?

What fields would be stored in a book record in the library database?

**IN – TEXT QUESTIONS**

- Enumerate five different types of files?
- Explain FIVE types of operations that can be performed on files with the help of suitable examples.
- Differentiate between a field and a record.

## 4.0 CONCLUSION

You have learnt of different files, their operations and how they organized in order to use them effectively. We have relative and sequential files based on the ease of accessibility of records store in them. Certain operations are carried out on files in form of opening, reading, updating, committing and closing. Most computers organize files into hierarchies using *folders*, *directories*, or *catalogs*. Each folder can contain an arbitrary number of files, and it can also contain other folders. These other folders are referred to as *subfolders*. Subfolders can contain still more files and folders and so on, thus building a tree-like structure in which one "master folder" (or "root folder" — the name varies from one operating system to another) can contain any number of levels of other folders and files. You could save your assignment and project data in different files for you to make use of them.

## 5.0 SUMMARY

You covered file management and organization in this unit. Files are organized sequentially, relatively or indexed. Most computer files are stored on hard disks and also backed up in flash drives and CDS.

## 6.0 TUTOR-MARKED ASSIGNMENT

1. Define the term File.
2. Explain the architecture of file organization.
3. What do you mean by field, record and table?
4. Define the term Data Capturing. Explain different data capturing techniques.
5. Explain what is meant by the term back–up? Why it is important to keep the back-up copy away from the computer system?
6. Validation and Verification help to reduce the errors when inputting data. Justify the statement.

7. Explain the difference between validation and verification. Give the names of three validations checks that can be used.

## 7.0 REFERENCES/SUGGESTED READINGS

Sinha, P.K. (2015). *Computer Fundamental.* Washington, D.C.: BPB Publications

Walmsley, D. (2014). *Information and Communication Technology.* Scotland: Hodder Murray 2

# MODULE 2: DATA PROCESSING IN EDUCATIONAL RESEARCH



## UNIT 1: RESEARCH PROBLEM: IDENTIFICATION AND FORMULATION

**CONTENTS**



### 1.0 INTRODUCTION

Research comprises creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications. Research problems are questions that indicate gaps in the

scope or the certainty of our knowledge. They point either to problematic phenomena, observed events that are puzzling in terms of our currently accepted ideas, or to problematic theories, current ideas that are challenged by new hypotheses.

This unit, the reader will understand the characteristics of a research problem or phenomenon.

understand the characteristics of good research questions.

## 2.0 OBJECTIVES

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

i. define the term research problem

ii. mention two roles of theory in research formulation

iii. list sources of research problem

iv. explain both internal and external criteria for research problem formulation

v. describe considerations in selecting a research problem

vi. discuss characteristics of a good research problem.

## 3.0 CONTENTS

### 3.1 What is a research problem?

It is the topic we would like to address, investigate, or study, whether descriptively or

experimentally. It is the focus or reason for engaging in our research. It is typically a topic, phenomenon, or challenge that we are interested in and with which we are at least somewhat familiar applications. A research problem, or phenomenon as it might be called in many forms of qualitative research, is the topic you would like to address, investigate, or study, whether descriptively or experimentally. It is the focus or reason for engaging in your research. It is typically a topic, phenomenon, or challenge that you are interested in and with which you are at least somewhat familiar.

### 3.2 Problem Formulation

#### 3.2.1 Concept of research formulation

We often think we understand problems when we don't. For example, when students encounter difficulties with word problems in math, teachers may initially think that students have not mastered the basic skills that would allow them to carry out the needed computations. However, the difficulty may actually lie in poor reading skills, which prevent the students from identifying the words in math problems. As another example, when students do not hand in homework assignments or participate in class, some might be inclined to think that the students are not motivated. While there may be motivational issues, motivation may not be the only factor. A high school student may have an evening job that demands considerable time

and energy. A younger student may be trying desperately to camouflage poor or nonexistent skills. In some cases, the chosen instructional strategy may not be well matched to the student's cognitive or attention level. Therefore, it is crucial that researchers accurately identify the problem they want to study.

Problem formulation is the logical first step toward this goal. As Northrop (1966) writes, "Inquiry starts only when something is unsatisfactory, when traditional beliefs are inadequate or in question, when the facts necessary, to resolve one's uncertainties are not known, when the most likely relevant hypotheses are not even imagined. What one has at the beginning of the inquiry is merely the problem.

The formulation of research problems also has an important social function. As Merton, Broom, and Cottrell (1959) suggest, researchers must justify the demands for attention and other scarce resources that research makes: "In conferring upon the scientist the right to claim that a question deserves the concerted attention of others as well as himself, the social institution of science exacts the obligation that he justify the claim.

Achieving significant research results is perhaps the most powerful justification for such claims, but this type of justification can be offered only after the fact, and only in the event that the research is successful. A compelling research problem, by contrast, must marshal support in advance of research and, if it is sufficiently compelling, can even sustain that support through the sometimes fruitless periods that researcher's experience. However, despite research problems' logical priority in inquiry, and their importance as a priori justification, a problem formulation, as John Dewey stresses, is in fact a "progressive" matter.

### 3.2.2 The Role of Theory in Problem Formulation

Theory plays a dual role in research.

1. On the one hand, new theories solve research problems by Research Problem: accounting for unexplained phenomena and by superseding questionable older theories.

2. On the other hand, existing theory guides researchers in formulating research problems. In determining whether and in what respects a phenomenon or a theory is problematic, researchers consider the context of accumulated theoretical as well as empirical knowledge.

## 3.3 Sources of Research Problem

< Classroom < School < Community

< Own-teaching experiences < Classroom lectures < Class discussions

< Seminars/workshops/paper presentations < Internet

< Out-of-class exchange of ideas with fellow students and professors

| | | | | | |
|---|---|---|---|---|---|
| < | Reading assignment | < | Textbook | < | Special programme |
| < | Research reports | < | Term papers | < | Consultation with |
| < | Course instructor | < | Advisor | < | Faculty member |

Other sources include

The research problem may be selected from the following sources:

< Theory of one's own interest

< Daily problems

< Technological changes

< Unexplored areas

< Discussions with other people

### 3.4 Criteria for Formulating the Problem

The selection of one appropriate researchable problem out of the identified problems requires evaluation of those alternatives against certain criteria, which may be grouped into:

#### 3.4.1 Internal Criteria

Internal Criteria consists of:

1. Researcher's interest: The problem should interest the researcher and be a challenge to him. Without interest and curiosity, he may not develop sustained perseverance Interest in a problem depends upon the researcher's educational background, experience, outlook and sensitivity.
2. Researcher's own resource: In the case of a research to be done by a researcher on his own, consideration of his own financial resource is pertinent. If it is beyond his means, he will not be able to complete the work, unless he gets some external financial support. Time resource is more important than finance. Research is a time-consuming process; hence it should be properly utilized.
3. Researcher's competence: A mere interest in a problem will not do. The researcher must be competent to plan and carry out a study of the problem. He must possess adequate knowledge of the subject-matter, relevant methodology and statistical

   procedures.

#### 3.4.2 External Criteria

1. Research-ability of the problem: The problem should be researchable, i.e., amendable for

finding answers to the questions involved in it through the scientific method.

2. Novelty of the problem: The problem must have novelty. There is no use of wasting one's time and energy on a problem already studied thoroughly by others.

3. Importance and urgency: Problems requiring investigation are unlimited, but available research efforts are very much limited.

4. Facilities: Research requires certain facilities such, as well equipped library facility, suitable and competent guidance, data analysis facility, etc. Hence the availability of the facilities relevant to the problem must be considered. Problems for research, their relative importance and significance should be considered.

5. Feasibility: A problem may be a new one and also important, but if research on it is not feasible, it cannot be selected.

6. Usefulness and social relevance: Above all, the study of the problem should make a

significant contribution to the concerned body of knowledge or to the solution of some significant practical problem. It should be socially relevant.

7. Research personnel: Research undertaken by professors and by research rganizations require the services of investigators and research officers. But in India and other developing countries, research has not yet become a prospective profession. Hence, talent persons are not attracted to research projects. Each identified problem must be evaluated in terms of the above internal and external criteria and the most appropriate, one may be selected by a research scholar.

### 3.5 Considerations in Selecting a Research Problem

These help to ensure that your study will remain manageable and that you will remain motivated.

1. Interest: a research endeavour is usually time consuming, and involves hard work and possibly unforeseen problems. One should select topic of great interest to sustain the required motivation.

2. Magnitude: It is extremely important to select a topic that you can manage within the time and resources at your disposal. Narrow the topic down to something manageable, specific and clear.

3. Level of expertise: Make sure that you have an adequate level of expertise for the task you are proposing since you need to do the work yourself.

4. Relevance: Ensure that your study adds to the existing body of knowledge, bridges

current gaps and is useful in policy formulation. This will help you to sustain interest in the study.

5. Availability of data: Before finalizing the topic, make sure that data are available.

6. Ethical issues: How ethical issues can affect the study and how ethical problemscan be

overcome should be thoroughly examined at the problem formulating stage.

### 3.6 Principle Components in the Formulation of a Problem

< The originating questions (what one wants to know?)

< The rational- theoretical or practical (why one wants to have the questions answered?)

< The specifying questions (possible answers to the originating) questions in term of that satisfy the rationale.)

The Originating Questions

< Represent the beginning of certain difficulties or challenges

< Are formulated in such specific indicate where exactly the answers to them can be searched for.

< Constitute the initial phase in the process of problem formulation.

< May be formulated in terms of broader delimited categories of social variable but do not indicate specifically which particular variables in each class might be germane to the issues.

< Usually derive from a general theoretical orientation rather than a definite theory.

Rationale of Questions

< Is the statement of reasons why a particular question is worth putting across.

< States what will happen to other parts of knowledge or practice if the question posed is answered, i.e., how the answer to the question will contribute to theory and/ or

practice.

< Helps to effect discrimination between scientifically good and scientifically trivial questions.

Specifying Questions

< Culminate the process of formulating a research problem

< Involve the breaking down of originating question in with several specifying questions related to particular aspects and their consequences.

### 3.7 Characteristics of a Good Research Problem

The following are detailed list of criteria for the choice of research problem.

Novelty-It should be sufficiently original so that it does not involve objectionable duplication.

Interesting-The problem should be interesting for the investigator himself.

Importance-If it is not worth-while, if adds to neither knowledge nor lead to any improvements in the current practices.

Immediate Application-The investigator should ask himself the question, will my research help in solving an urgent problem

Feasibility or Amenability-Feasibility issue of research includes the following

< Availability of data

< Availability of cooperation

< Availability of guidance

< Availability of other facilitates

< Experience and creativity

< Coverage and confidence

Common Errors in Formulating Research Problem

- Naming a Broad Filed to choose the broad area of study instead of specific problem makes no justification.
- Narrowing or Localizing a Topic
- The problem should not be narrowed to such an extent that it becomes too small and insignificant from research point or view.

## IN-TEXT QUESTIONS

Now that you have finished studying this unit, attempt these questions to revise what you have learnt

- Define research problem
- Mention five sources of research problem
- List three internal criteria for research problem
- Explain four areas of feasibility issue in research problem formulation

## 4.0 CONCLUSION

In concluding lines we need to say that research is a creative work and this unit gave the researcher an outline about the identification and formulation of research problems. Best selected problems serve its purpose and prove the solution to many identified and unidentified problems.

## 5.0 SUMMARY

This unit exposed you to intricacies involved in research problem formulation. We also looked at both internal and external criteria for generating research problem. Sources for research problem includes schools, literatures, experiences, Special programme, Research reports, term papers, consultation with, course instructor, advisor, faculty member among others.

## 6.0 TUTOR-MARKED ASSIGNMENT

i. Explain both internal and external criteria for research problem formulation ii. Discuss considerations in selecting a research problem
iii. Discuss characteristics of a good research problem.

## 7.0 REFERENCES / FURTHER READINGS


Ali, A. (2013). Approaches to Educational Research: analysis, criticisms and relevance. Perspectives in Educational Research and National Development (Vol. 1) Onitsha, Summer Educational Publishers Limited

Gall, M. D., Gall, J. P., & Borg, W. R. (2014). Educational research: An introduction (8th Ed.). Boston: Allyn & Bacon

Kothari, C.R. (2016). Research methodology: Methods & techniques. India: New Age International, Publishers

Koul, L. (2014). Methodology of educational research. India: Vikas Publishing House PVT Ltd

## UNIT 2: RESEARCH QUESTIONS AND HYPOTHESES FOR DATA PROCESSING

## CONTENTS



### 1.0 INTRODUCTION

Long before you create a research proposal, let alone conduct your research, you need to identify a problem to address and then a question or questions to ask regarding your targeted problem. This chapter first discusses the nature of a research problem, where you might get ideas for a problem to investigate, narrowing down or focusing on a particular problem to address, and writing good research questions. It then discusses finding literature that is relevant to and helpful in clarifying your targeted problem and question(s).

### 2.0 OBJECTIVES

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

i. define a research question
ii. explain PICO approach to research question
iii. explain the characteristics of research questions
iv. describe hypotheses with specific example

### 3.0 MAIN CONTENTS

### 3.1 DEFINITION OF RESEARCH QUESTION

A research question is a way of expressing your interest in a problem or phenomenon. Research questions are not necessarily an attempt to answer the many philosophical questions that often arise in schools, and they are certainly not intended to be an avenue for grinding personal axes regarding classroom or school issues. You may have more than one research question for a study, depending on the complexity and breadth of your proposed work. Each question should be clear and specific, refer to the problem or phenomenon, reflect. Questions that guide your research. Ideally, a research question should be debatable and of interest to both you and your potential readers.

It should also be based on a narrow topic. For instance, if you began your research with a broad, general interest in rehabilitation from Stroke, you might narrow your focus enough to ask the research question, “Does stroke rehabilitation (i.e., Physical Therapy, Occupational Therapy, Speech and Language Therapy) influence the psycho-social aspects of recovery?"
Whatever form the question takes, it needs to be well-defined. One useful way of focusing a research question is to use the PICO approach:
People or population- who are you asking the question about?
Intervention- what intervention are you interested in?
Control or comparison- what are you comparing the intervention to?
Outcome- what outcome are you interested in measuring?
Although this approach may only seem relevant to experimental research, with some minor modification it can be applied to studies of causation, diagnostic tests or prediction.

## 3.2 CHARACTERISTICS OF RESEARCH QUESTIONS

For research question to be complete and embracing it has to have the following features:

- FINER

  Feasible
  Adequate numbers of subjects?
  Adequate technical expertise?
  Affordable in time and money?
  Is it possible to measure or manipulate the variables?
  Interesting to the investigator?
  Novel to the field?

- ETHICAL

  Potential harm to subjects?
  Potential breech of subject confidentiality?
- Relevant to scientific knowledge/theory, organizational, health or social management and policy and to individual welfare.

## 3.3 CONCEPT OF HYPOTHESES

A research hypothesis is a statement of expectation or prediction that will be tested by research. Before formulating your research hypothesis, read about the topic of interest to you. From your reading, which may include articles, books and/or cases, you should gain sufficient information about your topic that will enable you to narrow or limit it and express it as a research question. The research question flows from the topic that you are considering. The research question, when stated as one sentence, is your Research Hypothesis.
In some disciplines, the hypothesis is called a “thesis statement.” Other words for “hypothesized” are “posited,” “theorized” or “proposed”. Remember, your hypothesis must REQUIRE two or more disciplines, one of which is law. This is essential, since your paper is interdisciplinary and a demonstration of the interdisciplinary process.
In your hypothesis, you are predicting the relationship between variables. Through the disciplinary insights gained in the research process throughout the year, you “prove” your

hypothesis. This is a process of discovery to create greater understandings or conclusions. It is not a strict proof as in logic or mathematics.

Following are some hints for the formulation of your hypothesis:

1. Be sure to read on the topic to familiarize yourself with it before making a final decision. You need to make certain that the topic is researchable in an interdisciplinary sense, meaning that there is sufficient published material on the topic in the legal literature and in the literature of a second or possibly a third discipline to write a 35-page paper.
2. As noted, a research hypothesis is more than just a topic. It has two elements (variables) that are in relation to each other. Remember that, within the word "hypothesis" is the word "thesis." Your hypothesis is what you propose to "prove" by your research. As a result of your research, you will arrive at a conclusion, a theory, or understanding that will be useful or applicable beyond the research itself.
3. Avoid judgmental words in your hypothesis. Value judgments are subjective and are not appropriate for a hypothesis. You should strive to be objective. Therefore the use of personal opinion is to be avoided.
4. Your hypothesis must involve an issue or question that cannot be answered exclusively by the discipline of law. You should try to limit your inquiry to the literatures of 2 or 3 disciplines. It is best to choose a hypothesis where you already have some level of familiarity with the disciplines that are most relevant to the topic.
5. Be sure that each term in your hypothesis is clearly understood and defined; do not deal in generalities or assume that the reader knows the meaning of a technical term.
6. Specify, if appropriate, whether you will be dealing with state or federal law or both on a comparative basis if appropriate.
7. Know that your hypothesis may change over time as your research progresses. You must obtain the professor's approval of your hypothesis, as well as any modifications to your hypothesis, before proceeding with any work on the topic.
   Your will be expressing your hypothesis in 3 ways:
   As a one-sentence hypothesis
   As a research question
   As a title for your paper

Your hypothesis will become part of your research proposal.
Note how each student, in the samples below, began with a general topic or area of interest, which evolved into a hypothesis. Look for the variables that each student subsequently explored in his/her paper. The examples below are final form hypotheses, which have been revised throughout the research process. You will find that your hypothesis may undergo changes too, as your research progresses.
Examples:

i. There is no significant difference in achievement between male and female Business Education students
ii. There is no significant relationship between students' attitudes and academic performance in research methods
iii. There is no significant main effect of socio-economic background of students' motivation to mathematics

## 3.4 TYPES OF HYPOTHESES

There are two types of hypotheses: Null hypothesis and Alternate hypothesis.

### 3.4.1 Null Hypothesis

This deals with no difference between parameters being measured. The following are examples of null hypotheses:

i. There is no significant difference in attitudes towards Business Education between male and female distance education students

ii. There is no significant relationship between students' reading habit and academic performance in Research methods

iii. There is no significant main effect of employment commitment on ODL students' class of degree after convocation.

Null hypothesis is denoted by Ho

### 3.4.2 Alternative Hypothesis

Alternative hypothesis states categorically the direction of difference. It is therefore referred to as directional hypothesis. Examples are

Attitudes of female distance learning students to Business Education is more favourable than their male counterparts

The relationship between students' reading habit and academic performance is negative

There is no significant main effect of employment commitment on ODL students' class of degree after convocation.

Null hypothesis is denoted by $H_1$.

You need to note that in educational research, hypotheses are tested at 0.05 level of significant.

## 3.5 CHARACTERISTICS OF A HYPOTHESIS

Oredein (2004) enumerated the following as the features of a good hypothesis:

i. A hypothesis should be free from vagueness and ambiguity

ii. It should be formulated in precise terms and consistent with current knowledge

iii. It should be critically and logically evaluated

iv. It should be statement of an expected relationship between two or more variables

v. It should be plausible; that is, it should be able to provide relevant solutions to the problem at hand

## 3.6 FUNCTIONS OF A HYPOTHESIS

i. Hypothesis provides the necessary structure for meaningful interpretation of data in relation to the problem under investigation

ii. It helps the researcher to understand the problem better in all ramifications as well as to identify the methods and procedures involved

iii. It provides a framework for drawing conclusion

iv. It sensitizes and draws the attention of the researcher to the important aspects of the problem under investigation

v. It gives a guide and direction to the investigation and solution for the problem at hand.

## IN-TEXT QUESTIONS

You can now reflect on what you have studied

- A research question should be debatable and of interest to both you and your potential readers. True / False
- Distinguish between research question and hypothesis
- Mention two types of hypothesis

### 4.0 CONCLUSION

After you have identified your research problem, the next step is to derive your research questions and generate hypotheses. This unit had explained what research questions are with various considerations you need to take to form them. Also, hypotheses are generated in form of null or alternative. Hypotheses are met to be plausible, testable and not ambiguous.

## 5.0 SUMMARY

Research questions are not necessarily an attempt to answer the many philosophical questions that often arise in schools, and they are certainly not intended to be an avenue for grinding personal axes regarding classroom or school issues. The research question flows from the topic that you are considering. The research question, when stated as one sentence, is your Research Hypothesis. Null and alternative hypotheses existed but in educational research it is advisable to state hypotheses in null form.

## 6.0 TUTOR-MARKED ASSIGNMENT

1. What is the difference between a research problem and a research question?
2. What makes a good research question?
3. What's the difference between a research question and a hypothesis?
4. Is this a good research question? Why or why not?

*Does peer tutoring affect the performance of ESL students on essays written in language arts classes?*

## 7.0 REFERENCES/ FURTHER READINGS

Alina, A. (2016). Statistics for Social Science and Public Policy. Ondo: Alex Publishers

Oredein, A. (2014). *Research Methods.* Lagos: Ababa Press Ltd

# UNIT 3 RESEARCH APPROACHES AND DESIGNS I

## CONTENTS



## 3.1 RESEARCH APPROACHES

### 3.1.1 Quantitative Research

Quantitative research is generally associated with the positivist/postpositivist paradigm. It usually involves collecting and converting data into numerical form so that statistical calculations can be made and conclusions drawn. Quantitative research is a type of research in which the researcher studies a problem that calls for an explanation about variables; decides what to study; asks specific, narrow questions; collects quantifiable data from participants; analyzes these numbers using statistics and graphs; and conducts the inquiry in an unbiased, objective manner (Clark & Creswell, 2015).

**The process**

Researchers will have one or more hypotheses. These are the questions that they want to address which include predictions about possible relationships between the things they want to investigate (variables). In order to find answers to these questions, the researchers will also have various instruments and materials (e.g. paper or computer tests, observation check lists etc.) and a clearly defined plan of action. Data is collected by various means following a strict procedure and prepared for statistical analysis. Nowadays, this is carried out with the aid of sophisticated statistical computer packages. The analysis enables the researchers to determine to what extent there is a relationship between two or more variables. This could be a simple association (e.g. people who exercise on a daily basis have lower blood pressure) or a causal relationship (e.g. daily exercise

actually leads to lower blood pressure). Statistical analysis permits researchers to discover complex causal relationships and to determine to what extent one variable influences another.

The results of statistical analyses are presented in journals in a standard way, the end result being a P value. For people who are not familiar with scientific research jargon, the discussion sections at the end of articles in peer reviewed journals usually describe the results of the study and explain the implications of the findings in straightforward terms

**Principles**

Objectivity is very important in quantitative research. Consequently, researchers take great care to avoid their own presence, behaviour or attitude affecting the results (e.g. by changing the situation being studied or causing participants to behave differently). They also critically examine their methods and conclusions for any possible bias.

Researchers go to great lengths to ensure that they are really measuring what they claim to be measuring. For example, if the study is about whether background music has a positive impact on restlessness in residents in a nursing home, the researchers must be clear about what kind of music to include, the volume of the music, what they mean by restlessness, how to measure restlessness and what is considered a positive impact. This must all be considered, prepared and controlled in advance.

External factors, which might affect the results, must also be controlled for. In the above example, it would be important to make sure that the introduction of the music was not accompanied by other changes (e.g. the person who brings the CD player chatting with the residents after the music session) as it might be the other factor which produces the results (i.e. the social contact and not the music). Some possible contributing factors cannot always be ruled out but should be acknowledged by the researchers.

The main emphasis of quantitative research is on deductive reasoning which tends to move from the general to the specific. This is sometimes referred to as a top down approach. The validity of conclusions is shown to be dependent on one or more premises (prior statements, findings or conditions) being valid. Aristotle's famous example of deductive reasoning was: All men are mortal à Socrates is a man à Socrates is mortal. If the premises of an argument are inaccurate, then the argument is inaccurate. This type of reasoning is often also associated with the fictitious character Sherlock Holmes. However, most studies also include an element of inductive reasoning at some stage of the research (see section on qualitative research for more details).

Researchers rarely have access to all the members of a particular group (e.g. all people with dementia, carers or healthcare professionals). However, they are usually interested in being able to make inferences from their study about these larger groups. For this reason, it is important that the people involved in the study are a representative sample of the wider population/group. However, the extent to which generalizations are possible depends to a certain extent on the

number of people involved in the study, how they were selected and whether they are representative of the wider group. For example, generalizations about psychiatrists should be based on a study involving psychiatrists and not one based on psychology students. In most cases, random samples are preferred (so that each potential participant has an equal chance of participating) but sometimes researchers might want to ensure that they include a certain number of people with specific characteristics and this would not be possible using random sampling methods. Generalizability of the results is not limited to groups of people but also to situations. It is presumed that the results of a laboratory experiment reflect the real-life situation which the study seeks to clarify.

When looking at results, the P value is important. P stands for probability. It measures the likelihood that a particular finding or observed difference is due to chance. The P value is between 0 and 1. The closer the result is to 0, the less likely it is that the observed difference is due to chance. The closer the result is to 1, the greater the likelihood that the finding is due to chance (random variation) and that there is no difference between the groups/variables.

### 3.1.2 Qualitative research

Qualitative research is the approach usually associated with the social constructivist paradigm which emphasises the socially constructed nature of reality. It is about recording, analysing and attempting to uncover the deeper meaning and significance of human behaviour and experience, including contradictory beliefs, behaviours and emotions. Researchers are interested in gaining a rich and complex understanding of people's experience and not in obtaining information which can be generalized to other larger groups. Qualitative research reports contribute new knowledge about the meaning of a phenomenon, such as what it means to be bullied as a teenage girl or what it means to survive a violent event. Qualitative research reports are able to provide in-depth, detailed descriptions of phenomena (Clark & Creswell, 2015).

**The process**

The approach adopted by qualitative researchers tends to be inductive which means that they develop a theory or look for a pattern of meaning on the basis of the data that they have collected. This involves a move from the specific to the general and is sometimes called a bottom-up approach. However, most research projects also involve a certain degree of deductive reasoning (see section on quantitative research for more details).

Qualitative researchers do not base their research on pre-determined hypotheses. Nevertheless, they clearly identify a problem or topic that they want to explore and may be guided by a theoretical lens - a kind of overarching theory which provides a framework for their investigation. The approach to data collection and analysis is methodical but allows for greater flexibility than in quantitative research. Data is collected in textual form on the basis of observation and interaction with the participants e.g. through participant observation, in-depth interviews and focus groups. It is not converted into numerical form and is not statistically analysed.

Data collection may be carried out in several stages rather than once and for all. The researchers may even adapt the process mid-way, deciding to address additional issues or dropping questions which are not appropriate on the basis of what they learn during the process. In some cases, the researchers will interview or observe a set number of people. In other cases, the process of data collection and analysis may continue until the researchers find that no new issues are emerging.

**Principles**

Researchers will tend to use methods which give participants a certain degree of freedom and permit spontaneity rather than forcing them to select from a set of pre-determined responses (of which none might be appropriate or accurately describe the participant's thoughts, feelings, attitudes or behaviour) and to try to create the right atmosphere to enable people to express themselves. This may mean adopting a less formal and less rigid approach than that used in quantitative research.

It is believed that people are constantly trying to attribute meaning to their experience. Therefore, it would make no sense to limit the study to the researcher's view or understanding of the situation and expect to learn something new about the experience of the participants. Consequently, the methods used may be more open-ended, less narrow and more exploratory (particularly when very little is known about a particular subject). The researchers are free to go beyond the initial response that the participant gives and to ask why, how, in what way etc. In this way, subsequent questions can be tailored to the responses just given. Qualitative research often involves a smaller number of participants. This may be because the methods used such as in-depth interviews are time and labour intensive but also because a large number of people are not needed for the purposes of statistical analysis or to make generalizations from the results.

The smaller number of people typically involved in qualitative research studies and the greater degree of flexibility does not make the study in any way "less scientific" than a typical quantitative study involving more subjects and carried out in a much more rigid manner. The objectives of the two types of research and their underlying philosophical assumptions are simply different. However, as discussed in the section on "philosophies guiding research", this does not mean that the two approaches cannot be used in the same study.

### 3.1.3 Pragmatic approach to research (mixed methods)

The pragmatic approach to science involves using the method which appears best suited to the research problem and not getting caught up in philosophical debates about which is the best approach. Pragmatic researchers therefore grant themselves the freedom to use any of the methods, techniques and procedures typically associated with quantitative or qualitative research. They recognise that every method has its limitations and that the different approaches can be complementary. Mixed methods involve the collection and "mixing" or integration of both quantitative and qualitative data in a study. Mixed methods research resides in the middle of this continuum because it incorporates elements of both qualitative and quantitative approaches (Creswell, 2014).

They may also use different techniques at the same time or one after the other. For example, they might start with face-to-face interviews with several people or have a focus group and then use the findings to construct a questionnaire to measure attitudes in a large-scale sample with the aim of carrying out statistical analysis.

Depending on which measures have been used, the data collected is analysed in the appropriate manner. However, it is sometimes possible to transform qualitative data into quantitative data and vice versa although transforming quantitative data into qualitative data is not very common.

Being able to mix different approaches has the advantages of enabling triangulation. Triangulation is a common feature of mixed methods studies. It involves, for example:

The use of a variety of data sources (data triangulation)

The use of several different researchers (investigator triangulation)

The use of multiple perspectives to interpret the results (theory triangulation)

The use of multiple methods to study a research problem (methodological triangulation)

In some studies, qualitative and quantitative methods are used simultaneously. In others, first one approach is used and then the next, with the second part of the study perhaps expanding on the results of the first. For example, a qualitative study involving in-depth interviews or focus group discussions might serve to obtain information which will then be used to contribute towards the development of an experimental measure or attitude scale, the results of which will be analysed statistically.

### 3.1.4 Advocacy/participatory approach to research (emancipatory)

To some degree, researchers adopting an advocacy/participatory approach feel that the approaches to research described so far do not respond to the needs or situation of people from marginalised or vulnerable groups. As they aim to bring about positive change in the lives of the research subjects, their approach is sometimes described as emancipatory. It is not a neutral stance. The researchers are likely to have a political agenda and to try to give the groups they are studying a voice. As they want their research to directly or indirectly result in some kind of reform, it is important that they involve the group being studied in the research, preferably at all *stages*, so as to avoid further marginalising them.

The researchers may adopt a less neutral position than that which is usually required in scientific research. This might involve interacting informally or even living amongst the research participants (who are sometimes referred to as co-researchers in recognition that the study is not simply about them but also by them). The findings of the research might be reported in more personal terms, often using the precise words of the research participants. Whilst this type of research could by criticised for not being objective, it should be noted that for some groups of people or for certain situations, it is necessary as otherwise the thoughts, feelings or behaviour of

the various members of the group could not be accessed or fully understood. Vulnerable groups are rarely in a position of power within society. For this reason, researchers are sometimes members of the group they are studying or have something in common with the members of the group.

## 3.2 Research Designs

The function of a research design is to ensure that the evidence obtained enables you to effectively address the research problem logically and as unambiguously as possible. In social sciences research, obtaining information relevant to the research problem generally entails specifying the type of evidence needed to test a theory, to evaluate a program, or to accurately describe and assess meaning related to an observable phenomenon. The research design provides a blueprint or plan for how researchers collect, analyze, and report their data in a study.

With this in mind, a common mistake made by researchers is that they begin their investigations far too early, before they have thought critically about what information is required to address the study's research questions. Without attending to these design issues beforehand, the overall research problem will not be adequately addressed and any conclusions drawn will risk being weak and unconvincing. As a consequence, the overall validity of the study will be undermined.

Given this, the length and complexity of describing research designs in your paper can vary considerably, but any well-developed design will achieve the following:

- Identify the research problem clearly and justify its selection, particularly in relation to any valid alternative designs that could have been used,
- Review and synthesize previously published literature associated with the problem,
- Clearly and explicitly specify hypotheses [i.e., research questions] central to the research problem,
- Effectively describe the data which will be necessary for an adequate testing of the hypotheses and explain how such data will be obtained, and
- Describe the methods of analysis to be applied to the data in determining whether or not the hypotheses are true or false.

### 3.2.1 Action Research Design

The essentials of action research design follow a characteristic cycle whereby initially an exploratory stance is adopted, where an understanding of a problem is developed and plans are made for some form of intervention strategy. Then the intervention is carried out (the "action" in Action Research) during which time, pertinent observations are collected in various forms. The new interventional strategies are carried out, and this cyclic process repeats, continuing until a sufficient understanding of (or a valid implementation solution for) the problem is achieved. The protocol is iterative or cyclical in nature and is intended to foster deeper understanding of a given situation, starting with conceptualizing and particularizing the problem and moving through several interventions and evaluations.

Strength of action research

This is a collaborative and adaptive research design that lends itself to use in work or community situations. Design focuses on pragmatic and solution-driven research outcomes rather than testing theories. When practitioners use action research, it has the potential to increase the amount they learn consciously from their experience; the action research cycle can be regarded as a learning cycle. Action research studies often have direct and obvious relevance in improving practice and advocating for change. There are no hidden controls or preemption of direction by the researcher.

**Weakness of action research**

- It is harder to do than conducting conventional research because the researcher takes on responsibilities of advocating for change as well as for researching the topic.
- Action research is much harder to write up because it is less likely that you can use a standard format to report your findings effectively [i.e., data is often in the form of stories or observation].
- Personal over-involvement of the researcher may bias research results.
- The cyclic nature of action research to achieve its twin outcomes of action (e.g. change) and research (e.g. understanding) is time-consuming and complex to conduct.
- Advocating for change requires buy-in from participants.

### 3.2.2 Case Study Design

**Definition and Purpose**

A case study is an in-depth study of a particular research problem rather than a sweeping statistical survey or comprehensive comparative inquiry. It is often used to narrow down a very broad field of research into one or a few easily researchable examples. The case study research design is also useful for testing whether a specific theory and model actually applies to phenomena in the real world. It is a useful design when not much is known about an issue or phenomenon.

**Strengths of Case Study Design**

- Approach excels at bringing us to an understanding of a complex issue through detailed contextual analysis of a limited number of events or conditions and their relationships.
- A researcher using a case study design can apply a variety of methodologies and rely on a variety of sources to investigate a research problem.
- Design can extend experience or add strength to what is already known through previous research.
- Social scientists, in particular, make wide use of this research design to examine contemporary real-life situations and provide the basis for the application of concepts and theories and the extension of methodologies.
- The design can provide detailed descriptions of specific and rare cases.

### Weaknesses of Case Study Design

- A single or small number of cases offers little basis for establishing reliability or to generalize the findings to a wider population of people, places, or things.

- Intense exposure to the study of a case may bias a researcher's interpretation of the findings.
- Design does not facilitate assessment of cause and effect relationships.
- Vital information may be missing, making the case hard to interpret.
- The case may not be representative or typical of the larger problem being investigated.
- If the criteria for selecting a case is because it represents a very unusual or unique phenomenon or problem for study, then your interpretation of the findings can only apply to that particular case.

### 3.2.3 Causal Design

**Definition and Purpose**

Causality studies may be thought of as understanding a phenomenon in terms of conditional statements in the form, "If X, then Y." This type of research is used to measure what impact a specific change will have on existing norms and assumptions. Most social scientists seek causal explanations that reflect tests of hypotheses. Causal effect (nomothetic perspective) occurs when variation in one phenomenon, an independent variable, leads to or results, on average, in variation in another phenomenon, the dependent variable.

**Conditions necessary for determining causality:**

Empirical association -- a valid conclusion is based on finding an association between the independent variable and the dependent variable.

Appropriate time order -- to conclude that causation was involved, one must see that cases were exposed to variation in the independent variable before variation in the dependent variable.

Non-spuriousness -- a relationship between two variables that is not due to variation in a third variable.

**Strengths of Causal Design**

- Causality research designs assist researchers in understanding why the world works the way it does through the process of proving a causal link between variables and by the process of eliminating other possibilities.
- Replication is possible.
- There is greater confidence the study has internal validity due to the systematic subject selection and equity of groups being compared.

**Weaknesses of Causal Design**

- Not all relationships are casual! The possibility always exists that, by sheer coincidence, two unrelated events appear to be related

- Conclusions about causal relationships are difficult to determine due to a variety of extraneous and confounding variables that exist in a social environment. This means causality can only be inferred, never proven.
- If two variables are correlated, the cause must come before the effect. However, even though two variables might be causally related, it can sometimes be difficult to determine which variable comes first and, therefore, to establish which variable is the actual cause and which the actual effect is.

### 3.2.4 Cohort Design

## Definition and Purpose

Often used in the medical sciences, but also found in the applied social sciences, a cohort study generally refers to a study conducted over a period of time involving members of a population which the subject or representative member comes from, and who are united by some commonality or similarity. Using a quantitative framework, a cohort study makes note of statistical occurrence within a specialized subgroup, united by same or similar characteristics that are relevant to the research problem being investigated, rather than studying statistical occurrence within the general population. Using a qualitative framework, cohort studies generally gather data using methods of observation. Cohorts can be either "open" or "closed."

Closed Cohort Studies [static populations, such as patients entered into a clinical trial] involve participants who enter into the study at one defining point in time and where it is presumed that no new participants can enter the cohort. Given this, the number of study participants remains constant (or can only decrease).

**Strengths of Cohort Design**

- The use of cohorts is often mandatory because a randomized control study may be unethical. For example, you cannot deliberately expose people to asbestos, you can only study its effects on those who have already been exposed. Research that measures risk factors often relies upon cohort designs.
- Because cohort studies measure potential causes before the outcome has occurred, they can demonstrate that these “causes” preceded the outcome, thereby avoiding the debate as to which is the cause and which is the effect.
- Cohort analysis is highly flexible and can provide insight into effects over time and related to a variety of different types of changes [e.g., social, cultural, political, economic etc.].
- Either original data or secondary data can be used in this design.

**Weaknesses of Cohort Design**

- In cases where a comparative analysis of two cohorts is made [e.g., studying the effects of one group exposed to asbestos and one that has not], a researcher cannot control for all

other factors that might differ between the two groups. These factors are known as confounding variables.

- Cohort studies can end up taking a long time to complete if the researcher must wait for the conditions of interest to develop within the group. This also increases the chance that key variables change during the course of the study, potentially impacting the validity of the findings.
- Due to the lack of randomization in the cohort design, its external validity is lower than that of study designs where the researcher randomly assigns participants.

### 3.2.5 Cross-Sectional Design

#### Definition and Purpose

Cross-sectional research designs have three distinctive features: no time dimension; a reliance on existing differences rather than change following intervention; and, groups are selected based on existing differences rather than random allocation. The cross-sectional design can only measure differences between or from among a variety of people, subjects, or phenomena rather than a process of change. As such, researchers using this design can only employ a relatively passive approach to making causal inferences based on findings.

#### Strengths of Cross-sectional Design

- Cross-sectional studies provide a clear 'snapshot' of the outcome and the characteristics associated with it, at a specific point in time.
- Unlike an experimental design, where there is an active intervention by the researcher to produce and measure change or to create differences, cross-sectional designs focus on studying and drawing inferences from existing differences between people, subjects, or phenomena.
- Entails collecting data *at* and *concerning* one point in time. While longitudinal studies involve taking multiple measures over an extended period of time, cross-sectional research is focused on finding relationships between variables at one moment in time.
- Groups identified for study are purposely selected based upon existing differences in the sample rather than seeking random sampling.
- Cross-section studies are capable of using data from a large number of subjects and, unlike observational studies, is not geographically bound.
- Can estimate prevalence of an outcome of interest because the sample is usually taken from the whole population.
- Because cross-sectional designs generally use survey techniques to gather data, they are relatively inexpensive and take up little time to conduct.

#### Weaknesses of Cross-sectional Design

- Finding people, subjects, or phenomena to study that are very similar except in one specific variable can be difficult.
- Results are static and time bound and, therefore, give no indication of a sequence of events or reveal historical or temporal contexts.
- Studies cannot be utilized to establish cause and effect relationships.
- This design only provides a snapshot of analysis so there is always the possibility that a study could have differing results if another time-frame had been chosen.
- There is no follow up to the findings.

## IN-TEXT QUESTIONS

Now that you've been introduced to research approaches and designs in this unit, ascertain the level of your attainment by proffer answers to the following questions

- Positivist paradigm resulted in _____ research
- Qualitative research approach is connected to social constructivist paradigm. True or False?
- What does P value measure?
- What is data triangulation?

## 4.0 CONCLUSION

You have discovered from this unit that the main emphasis of quantitative research is on deductive reasoning which tends to move from the general to the specific. This is sometimes referred to as a top down approach. In qualitative research, researchers are interested in gaining a rich and complex understanding of people's experience and not in obtaining information which can be generalized to other larger groups. It is also essential that without attending design issues beforehand, the overall research problem will not be adequately addressed and any conclusions drawn will risk being weak and unconvincing. As a consequence, the overall validity of the study will be undermined.

## 5.0 SUMMARY

This unit had exposed you to three major approaches of research, namely quantitative, qualitative and advocacy approaches. It is also discovered that research designs are essential as to be considered when one wants to undertake a research. Action research, case study, causal comparative, cohort and cross-sectional designs are applicable to educational and business related studies. Consider the strengths and weaknesses of any design that you want to employ to have a solid ground on which the work is based.

## 6.0 TUROR-MARKED ASSIGNMENT

- Distinguish between qualitative and quantitative research approaches.
- Mention five research designs you know. Explain two of them.
- Discuss the principles and processes involved in advocacy research.

## 7.0 REFERENCES / FURTHER READINGS


Clark, V. L. P., & Creswell, J. W. (2015). *Understanding Research: A Consumer's Guide* (J. W. Johnston, G. Gottfried, & C. Griscom Eds. 2nd ed.). Boston: Pearson Education, Inc.

Creswell, J. W. (2014). *Research Design : Qualitative, Quantitative, and Mixed Methods Approaches* (V. Knight, J. Young, & K. Koscielak Eds. 4th ed.). Thousand Oaks, California: SAGE Publications, Inc.

F.N. & Lee (2016) Foundations of Behavioural Research. Australia: Thomson Learning Inc.

Nwana, O.C. (2015). Introduction to Educational Research. Ibadan: Heinemann Educational Books (Nigeria) Limited

# UNIT 4: RESEARCH APPROACHES AND DESIGNS II

## CONTENTS



## 1.0 INTRODUCTION

In unit 4 we discussed approaches to research in form of quantitative, qualitative and mixed methods. It was emphasized that the approach to explore depends on the designs selected. Some of the designs you came across in that unit were action research, case study, casual, cohort and cross-sectional designs. You will learn of other research designs that are available for your use when you are undertaken your research project in this unit.

## 2.0 OBJECTIVES

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

- Explain descriptive survey design
- Enumerate four areas of strengths and weaknesses of experimental research design.
- Identify possible insights to goals which exploratory research that are intended to produce.
- Describe longitudinal research design

## 3.0 CONTENTS

### 3.1.1 Descriptive Survey Design

### Definition and Purpose

Descriptive research designs help provide answers to the questions of who, what, when, where, and how associated with a particular research problem; a descriptive study cannot conclusively ascertain answers

to why. Descriptive research is used to obtain information concerning the current status of the phenomena and to describe "what exists" with respect to variables or conditions in a situation.

**Strengths of Descriptive Survey**

- The subject is being observed in a completely natural and unchanged natural environment. True experiments, whilst giving analyzable data, often adversely influence the normal behavior of the subject [a.k.a., the Heisenberg effect whereby measurements of certain systems cannot be made without affecting the systems].
- Descriptive research is often used as a pre-cursor to more quantitative research designs with the general overview giving some valuable pointers as to what variables are worth testing quantitatively.
- If the limitations are understood, they can be a useful tool in developing a more focused study.
- Descriptive studies can yield rich data that lead to important recommendations in practice.
- Approach collects a large amount of data for detailed analysis.

**Weaknesses of Descriptive Survey**

- The results from a descriptive research cannot be used to discover a definitive answer or to disprove a hypothesis.
- Because descriptive designs often utilize observational methods [as opposed to quantitative methods], the results cannot be replicated.
- The descriptive function of research is heavily dependent on instrumentation for measurement and observation.

### 3.1.2 Experimental Design

#### Definition and Purpose

A blueprint of the procedure that enables the researcher to maintain control over all factors that may affect the result of an experiment. In doing this, the researcher attempts to determine or predict what may occur. Experimental research is often used where there is time priority in a causal relationship (cause precedes effect), there is consistency in a causal relationship (a cause will always lead to the same effect), and the magnitude of the correlation is great. The classic experimental design specifies an experimental group and a control group. The independent variable is administered to the experimental group and not to the control group, and both groups are measured on the same dependent variable. Subsequent experimental designs have used more groups and more measurements over longer periods. True experiments must have control, randomization, and manipulation.

**Strengths of Experimental Design**

- Experimental research allows the researcher to control the situation. In so doing, it allows researchers to answer the question, “What causes something to occur?”
- Permits the researcher to identify cause and effect relationships between variables and to

distinguish placebo effects from treatment effects.

- Experimental research designs support the ability to limit alternative explanations and to infer direct causal relationships in the study.
- Approach provides the highest level of evidence for single studies.

**Weaknesses of Experimental Design**

- The design is artificial, and results may not generalize well to the real world.
- The artificial settings of experiments may alter the behaviors or responses of participants.
- Experimental designs can be costly if special equipment or facilities are needed.
- Some research problems cannot be studied using an experiment because of ethical or technical reasons.
- Difficult to apply ethnographic and other qualitative methods to experimentally designed studies.

### 3.1.3 Exploratory Design

### Definition and Purpose

An exploratory design is conducted about a research problem when there are few or no earlier studies to refer to or rely upon to predict an outcome. The focus is on gaining insights and familiarity for later investigation or undertaken when research problems are in a preliminary stage of investigation. Exploratory designs are often used to establish an understanding of how best to proceed in studying an issue or what methodology would effectively apply to gathering information about the issue.

The goals of exploratory research are intended to produce the following possible insights:

- Familiarity with basic details, settings, and concerns.
- Well grounded picture of the situation being developed.
- Generation of new ideas and assumptions.
- Development of tentative theories or hypotheses.
- Determination about whether a study is feasible in the future.
- Issues get refined for more systematic investigation and formulation of new research questions.
- Direction for future research and techniques get developed.

**Strengths of Explorative Design**

- Design is a useful approach for gaining background information on a particular topic.
- Exploratory research is flexible and can address research questions of all types (what, why, how).
- Provides an opportunity to define new terms and clarify existing concepts.
- Exploratory research is often used to generate formal hypotheses and develop more precise research problems.
- In the policy arena or applied to practice, exploratory studies help establish research

priorities and where resources should be allocated.

**Weaknesses of Explorative Design**

- Exploratory research generally utilizes small sample sizes and, thus, findings are typically not generalizable to the population at large.
- The exploratory nature of the research inhibits an ability to make definitive conclusions about the findings. They provide insight but not definitive conclusions.
- The research process underpinning exploratory studies is flexible but often unstructured, leading to only tentative results that have limited value to decision-makers.
- Design lacks rigorous standards applied to methods of data gathering and analysis because one of the areas for exploration could be to determine what method or methodologies could best fit the research problem.

### 3.1.4 Historical Design

**Definition and Purpose**

The purpose of a historical research design is to collect, verify, and synthesize evidence from the past to establish facts that defend or refute a hypothesis. It uses secondary sources and a variety of primary documentary evidence, such as, diaries, official records, reports, archives, and non-textual information [maps, pictures, audio and visual recordings]. The limitation is that the sources must be both authentic and valid.

**Strengths of Historical Design**

- The historical research design is unobtrusive; the act of research does not affect the results of the study.
- The historical approach is well suited for trend analysis.
- Historical records can add important contextual background required to more fully understand and interpret a research problem.
- There is often no possibility of researcher-subject interaction that could affect the findings.
- Historical sources can be used over and over to study different research problems or to replicate a previous study.

**Weaknesses of Historical Design**

- The ability to fulfill the aims of your research are directly related to the amount and quality of documentation available to understand the research problem.
- Since historical research relies on data from the past, there is no way to manipulate it to control for contemporary contexts.
- Interpreting historical sources can be very time consuming.
- The sources of historical materials must be archived consistently to ensure access. This

may especially be challenging for digital or online-only sources.

- Original authors bring their own perspectives and biases to the interpretation of past events and these biases are more difficult to ascertain in historical resources.
- Due to the lack of control over external variables, historical research is very weak with regard to the demands of internal validity.
- It is rare that the entirety of historical documentation needed to fully address a research problem is available for interpretation, therefore, gaps need to be acknowledged.

### 3.1.5 Longitudinal Design

#### Definition and Purpose

A longitudinal study follows the same sample over time and makes repeated observations. For example, with longitudinal surveys, the same group of people is interviewed at regular intervals, enabling researchers to track changes over time and to relate them to variables that might explain why the changes occur. Longitudinal research designs describe patterns of change and help establish the direction and magnitude of causal relationships. Measurements are taken on each variable over two or more distinct time periods. This allows the researcher to measure change in variables over time. It is a type of observational study sometimes referred to as a panel study.

**Strengths of Longitudinal Design**

- Longitudinal data facilitate the analysis of the duration of a particular phenomenon.
- Enables survey researchers to get close to the kinds of causal explanations usually attainable only with experiments.
- The design permits the measurement of differences or change in a variable from one period to another [i.e., the description of patterns of change over time].
- Longitudinal studies facilitate the prediction of future outcomes based upon earlier factors.

**Weaknesses of Longitudinal Design**

- The data collection method may change over time.
- Maintaining the integrity of the original sample can be difficult over an extended period of time.
- It can be difficult to show more than one variable at a time.
- This design often needs qualitative research data to explain fluctuations in the results.
- A longitudinal research design assumes present trends will continue unchanged.
- It can take a long period of time to gather results.
- There is a need to have a large sample size and accurate sampling to reach representativeness.

### 3.1.6 Meta-Analysis Design

#### Definition and Purpose

Meta-analysis is an analytical methodology designed to systematically evaluate and summarize the results from a number of individual studies, thereby, increasing the overall sample size and the ability of the researcher to study effects of interest. The purpose is to not simply summarize existing knowledge, but to develop a new understanding of a research problem using synoptic reasoning. The main objectives of meta-analysis include analyzing differences in the results among studies and increasing the precision by which effects are estimated. A well-designed meta-analysis depends upon strict adherence to the criteria used for selecting studies and the availability of information in each study to properly analyze their findings. Lack of information can severely limit the type of analyses and conclusions that can be reached. In addition, the more dissimilarity there is in the results among individual studies [heterogeneity], the more difficult it is to justify interpretations that govern a valid synopsis of results.

- A meta-analysis needs to fulfill the following requirements to ensure the validity of your findings:

  - Clearly defined description of objectives, including precise definitions of the variables and outcomes that are being evaluated;
  - A well-reasoned and well-documented justification for identification and selection of the studies;
  - Assessment and explicit acknowledgment of any researcher bias in the identification and selection of those studies;
  - Description and evaluation of the degree of heterogeneity among the sample size of studies reviewed; and,
  - Justification of the techniques used to evaluate the studies.

**Strengths of Meta-analysis Design**

- Can be an effective strategy for determining gaps in the literature.
- Provides a means of reviewing research published about a particular topic over an extended period of time and from a variety of sources.
- Is useful in clarifying what policy actions can be justified on the basis of analyzing research results from multiple studies.
- Provides a method for overcoming small sample sizes in individual studies that previously may have had little relationship to each other.
- Can be used to generate new hypotheses or highlight research problems for future studies.

**Weaknesses of Meta-analysis Design**

- Small violations in defining the criteria used for content analysis can lead to difficult to interpret and/or meaningless findings.
- A large sample size can yield reliable, but not necessarily valid, results.
- A lack of uniformity regarding, for example, the type of literature reviewed, how methods are applied, and how findings are measured within the sample of studies you are analyzing, can make the process of synthesis difficult to perform.
- Depending on the sample size, the process of reviewing and synthesizing multple studies can be very time consuming.

### 3.1.7 Observational Design

**Definition and Purpose**

This type of research design draws a conclusion by comparing subjects against a control group, in cases where the researcher has no control over the experiment. There are two general types of observational designs. In direct observations, people know that you are watching them. Unobtrusive measures involve any method for studying behavior where individuals do not know they are being observed. An observational study allows a useful insight into a phenomenon and avoids the ethical and practical difficulties of setting up a large and cumbersome research project.

**Strengths of Observational Design**

- Observational studies are usually flexible and do not necessarily need to be structured around a hypothesis about what you expect to observe [data is emergent rather than pre-existing].
- The researcher is able to collect in-depth information about a particular behavior.
- Can reveal interrelationships among multifaceted dimensions of group interactions.
- You can generalize your results to real life situations.
- Observational research is useful for discovering what variables may be important before applying other methods like experiments.
- Observation research designs account for the complexity of group behaviors.

**Weaknesses of Observational Design**

- Reliability of data is low because seeing behaviors occur over and over again may be a time consuming task and are difficult to replicate.
- In observational research, findings may only reflect a unique sample population and, thus, cannot be generalized to other groups.
- There can be problems with bias as the researcher may only "see what they want to see."
- There is no possibility to determine "cause and effect" relationships since nothing is manipulated.
- Sources or subjects may not all be equally credible.
- Any group that is knowingly studied is altered to some degree by the presence of the researcher, therefore, potentially skewing any data collected.

## IN-TEXT QUESTIONS

Make a quick revision of various research designs discussed in this unit by proffer answers to the following questions

- _____ design is often used as a pre-cursor to quantitative research design
- Control group is an essential aspect of historical design. Yes or No?
- Mention 2 important variables in an experimental research.
- When several studies are systematically evaluated and summarized to develop new understanding involves _____ research design

## 4.0 CONCLUSION

As you have studied in this unit, research is carried out based on the design employ by the researcher. It is required that such researchers need to be conversant with the designs available so as to use the appropriate ones to undertake the research. Your understanding of research designs give you the in road to carry out your research project as a n education student.

## 5.0 SUMMARY

In this unit, descriptive survey design is used to obtain information concerning the current status of the phenomena and to describe "what exists" with respect to variables or conditions in a situation. Other research designs like experimental, meta-analysis, historical, explorative and longitudinal were discussed. Each of the designs is effective in its own right, notwithstanding, each also has strengths and limitations that you need to consider during research processes.

## 6.0 TUTOR-MARKED ASSIGNMENT

- Discuss key features of descriptive survey design
- Mention four advantages and four disadvantages of experimental research
- List five insights the goals of explorative research intended to produce
- Differentiate between historical and longitudinal designs.

## 7.0 REFERENCES / FURTHER READINGS


Anastas, J. W. (1999). Flexible Methods: Relational and Longitudinal Research. New York: Columbia Cooper,

Clark, V. L. P., & Creswell, J. W. (2015). *Understanding Research: A Consumer's Guide* (J. W. Johnston, G. Gottfried, & C. Griscom Eds. 2nd ed.). Boston: Pearson Education, Inc.

Creswell, J. W. (2014). *Research Design : Qualitative, Quantitative, and Mixed Methods Approaches* (V. Knight, J. Young, & K. Koscielak Eds. 4th ed.). Thousand Oaks, California: SAGE Publications, Inc.

De Vaus, D. A. (2001). *Research Design in Social Research*. London: SAGE.

Gall, M. (2007). *Educational Research: An Introduction*. Boston, MA: Pearson/Allyn and Bacon.

Maykut, P. S.(1994). *Beginning Qualitative Research: A Philosophic and Practical Guide*. Washington,

H., Larry V. H., & Jeffrey C. V. (2009). *The Handbook of Research Synthesis and Meta- Analysis*. New York: Russell Sage Foundation. University Press.

Reason, P. & Hilary B. (2015). *Handbook of Action Research: Participative Inquiry and Practice*. Thousand Oaks, CA: SAGE.

Stanford Encyclopedia of Philosophy (2013). *Metaphysics Research Lab*, CSLI, Stanford University. D.C.: Falmer Press

Rosenbaum, P. R. (2010). *Design of Observational Studies*. New York: Springer.

# UNIT 5: Data Collection Tools and Methods

## CONTENTS



## 1.0 INTRODUCTION

To grow your market, you need to assess its current performance, strengths, and weaknesses. Assessment involves research—collecting, analyzing, and interpreting data. There are a number of research methods and tools that can be used to collect data on the performance of your farmers market. Technology can be used as a source of data (audio, film, visual digital materials, and web-based information), as a data collection tool (e.g., videotaping, audiotaping, e-interviews, and digital pictures), and as an analysis tool. There are a number of different methods of collecting the data you need to assess your market's performance, evaluate your strengths and weaknesses and position relative to your competitors, and create a strategic marketing plan. This section provides a brief overview of those data collection methods and their appropriate use.

## 2.0 Objectives

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

## 3.0 CONTENTS

### 3.1 Anecdotal Records

Anecdotes are stories. Anecdotal data simply means the information you get from the stories you hear about the market from day to day—vendors' stories about outrageous complaints or happy customers, customers' stories about their favorite market experience when you run into them at the library or a meeting, stories you tell your family at the dinner table. Anecdotal data is a good starting point for more formal research. From interacting with customers, the manager, staff, and vendors probably have a fairly good idea of basic issues and questions to be asked. The manager should also have a picture of the issues that concern the market's vendors and possibly of issues regarding the market that may concern area business owners and neighbors. Such data, while not gathered or

interpreted systematically, can point to the questions that should be the focus of more formal efforts to evaluate market performance and research.

### 3.2 Customer and Vendor Feedback

If the market has a suggestion box or a process for handling written complaints, the records of these also can be useful for evaluating themes and trends in the strengths and weaknesses of the market. Such themes and trends can be a basis for more formal research and analysis.

### 3.3 Interviews

Interviews can be valuable sources of in-depth information about the preferences and evaluations of respondents. Interviews can range from the very formal, following a strict protocol with the same questions asked of every respondent, to very informal, with the interviewer following the respondent's comments and allowing information to emerge based on those comments. Most interviews will fall somewhere in the middle with a common set of questions and leeway to follow up on interesting comments.

The primary advantage of the interview as a data collection tool is the chance to get more in-depth information than is possible in a survey. However, interviews also have a number of disadvantages. First, they are far more labor intensive to conduct and interpret than surveys. Someone must ask the interview questions of each individual respondent and either tape record or take detailed notes of the answers. The personality of the interviewer can have a huge effect on the quality and completeness of responses. If interviews are recorded, they must be played back, transcribed, or both. Because responses will probably contain a great deal of variation among respondents, it can be difficult and time-consuming to find the themes that are useful for improving the market. All these considerations mean that interviews are only practical with a small number of respondents, unlike the much larger number that can be reached with a survey. The two in combination can improve on the results of either interviews or surveys alone, and we discuss how to combine methods for maximum effectiveness in the chapter on evaluating your market.

### 3.4 Focus Groups

Focus groups are interviews in which the researchers bring together a group of people who do not necessarily know each other but who share a characteristic that is relevant to the question to be answered. If, for example, you wish to learn why some people in the target population do not shop at your farmers market, you might assemble a focus group of people who have shopped at the market in the past but do not do so now, or a group of people in your target population who have never shopped at the market. An advantage to focus group interviews is that they allow people to build on each other's ideas, reflecting and listening and developing their opinions through interaction. Such interviews can yield richer, more thoughtful answers than individual interviews.

Disadvantages of focus groups are that they are difficult to organize and that they require some skill on the part of the interviewer to keep the group on track. Furthermore, because the answers are the result of interaction, they may vary highly from group to group and so can be difficult to interpret. A group can easily be "hijacked" by the strong opinions of a single individual. Trustworthy results depend on conducting several focus group interviews and looking for common themes in the responses.

### 3.5 Tasting Panels

Related to a focus group, a tasting panel is a group of people gathered to evaluate market products. You can use tasting panels to assess your product mix and quality.

### 3.6 Observation

Another useful source of data is observation with specific criteria or questions in mind. Market management might gather data by observing and recording the level of activity at the market at different hours— is the market busy early in the day with business tapering off later or is it steady?

Do vendors tend to have large amounts of product left or run out early? Observation, especially when combined with other sources of data, can be very useful in evaluating the market's performance. One method of observation involves counting—numbers of customers per half hour or the number of customers who fit certain categories, such as families with young children or seniors, for example. Especially when evaluating your position relative to competitors such as supermarkets, you can use observations and comparisons; for example, you may compare price, quality, or selection of a sample of produce.

Apart from business research, observation is also applicable to education. Observation of students is an important aspect of classroom activities. It is most effective when focused on skills, concepts, and attitudes. Making brief notes and keeping checklists when watching students' activities during practical sections like typing enable teachers to keep records of continuous progress and achievement. Observation helps researchers to create specific occasions for observation and practice focus- that is, they select what and whom to observe, and put all else in the background.

### 3.6.1 Observation technique and its characteristics.

In educational research Interviews and observation are the primary data collection tools. An observational technique may be operational defined as a process whereby individuals or groups of people are ''commissioned'' to watch and record the happenings of events or even study behavioural patterns in settings of interest. It has been the prevailing method of inquiring. Observation continues to characterize all research: experimental, descriptive and qualitative.

Characteristics of a good observation include:

1. A carefully planned observation, in which observer knows what they are looking for and what is relevant in a situation.
2. Observers are aware of the wholeness of what is observed.
3. Observers are objective.
4. Observers separate the facts from the interpretation of facts.
5. Observations are checked and verified, whenever possible by repetition or by comparison with those of other competent observers.
6. Observations are carefully and expertly recorded.

### 3.6.2 Types and Uses of Observation technique.

Observations may be direct or inferential. An observation is said to be direct when the observer is involved in the first hand experiences of the happenings of given situations. But an observation is described as inferential when researcher draws inferences on the basis of the observation report supplied by another person or group of persons.

According to Okpala et al, Yoloye (1997), observational data could be useful in the following;

- Measuring classroom process variables
- Measuring programme implementation
- Identifying difficulties in programme use
- Identifying changes introduced by teachers

- Indentifying typical instructional pathways

### 3.7 Document Analysis

A functioning market already has invaluable sources of data in the records kept. Much can be learned from records of vendor sales or percentages of revenue paid as fees to the market and other financial records. Such records give a picture of the market's busy and less busy times and seasons.

## 3.8 Questionnaire

A questionnaire is a series of relevant questions which are usually used to elicit information from respondent who are normally drawn from the target population of a given study. A questionnaire may contain either structure or unstructured questions. The structure question otherwise known as close questions are those item in a questionnaire in which alternative responses are provided by the researcher from which the respondent is to select from. Example of a structure item of a questionnaire

In a period of dwindling national economy, parent should be responsible for paying the school fees of their wards at all level of education.

Alternative Responses Provided

A. Strongly agree

B. Agree
C. Undecided
D. Disagree
E. Strongly disagree.

An unstructured question or item (or open-ended item) is that item or question in which pre-determine responses are not provided for respondent to choose from, Example of an unstructured item of a questionnaire:

Questions: What do you consider to be the implication of open and distant learning on the future of a national economy? .....

---

---

### 3.8.1 Basic Guidelines for Designing and Organizing a Questionnaire

There are four basic guidelines for designing and organizing a questionnaire. These include:

- Introductory section of a questionnaire:
  Educational research is to note that every questionnaire should have an introductory section which should give a description of what such investigation was all about.
- Ordering of Questionnaire items;
  in any form of questionnaire item prepare, the items should always begin with simple items and examples of how to respond to the simple items. It is advisable to begin Questionnaire items with simple questions or items before delving into difficult items, and as much as possible, related questions should follow one another. If Questionnaire items are categorized, items in the same category

should be formulated in such a way that similar items follow one after the other.

- Types of Questionnaire Require for a given study:
  Whenever an educational researcher decides to use a Questionnaire as a research tool, it is advisable to decides right from the beginning, which form of the Questionnaire items (structure or unstructured items)are to be used for the investigation under study.
- Used of Language to formulate Questionnaire items:
  Educational researcher are implored to write Questionnaire items in simple and unambiguous English or Language so that comprehension of the item does not constitute problems for the intended respondents

**IN-TEXT QUESTIONS**

- List four research instruments and explain two types of questionnaire.
- **Differentiate between focus group and tasting panel**
- **Enumerate four limitations of interviews as research tool.**

## 4.0 Conclusion

The most important consideration for collecting high-quality data and interpreting it accurately knows what questions you want to answer with the information you gather. Once you have identified those questions, you can choose the most appropriate methods for gathering and interpreting information to answer them. With a focused effort, you will likely find the information you need to create a successful research efforts.

## 5.0 SUMMARY

This unit covered areas of research tools / instruments for gathering data. We have looked at various data collection tools such as anecdotal records, customer vendor feedback, interviews (structured and unstructured), focus group, tasting panels, document analysis and questionnaires. All these instruments are available for you to as Business education students in your research endeavor and business field.

## 6.0 TUTOR MARKED ASSIGNMENT

- Explain anecdotal records in relation to data collection in research.
- Enumerate the basic guidelines for constructing a good questionnaire.

## 7.0 REFERENCES / FURTHER READINGS

Ary, D., Jacobs, L. C., & Sorensen, C. (2017). *Introduction to Research in Education*. Canada: Wadsworth, Cengage Learning.

Dey, I. (2016). Quantitative data analysis. London: Taylor and francis Group.

Donald H. & Robbert, G. (2016). Longitudinal data analysis. Chicago: Wiley & Sons

Kirk, R.E. (2015). Statistical: An Introduction. Forth Worth: TX Holt, Rinehart and Wiston.

# MODULE 3: STATISTICS IN EDUCATIONAL RESEARCH



## UNIT 1: INTRODUCTION TO STATISTICS



### 1.0 INTRODUCTION

In the previous modules/units, you worked through the different methods of collecting data in research. The question is; what do you do with this seemingly unmanageable bulk of data?

This question will take us to 'Data Analysis', which we shall describe "as the process of organizing and summarizing data in order to provide answers to the research questions or test hypotheses stated in the study". This process, most of the times, involves the use of statistical procedures to summarize and describe the characteristics of samples and populations of the study.

In this unit, we shall first look at the meaning of statistics, the types of statistics and organization of data.

### 2.0 OBJECTIVES

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

(i) define the concept statistics;
(ii) explain the types of statistics;

(iii) organize a set of scores under (a) sequencing, (b) frequency distribution table, (c) bar chart.

## 3.0 MAIN CONTENT

### 3.1 Meaning of Statistics

Statistics, as a word, has different shades of meaning. These meanings can be in the plural form or singular form.

**(i) It is regarded as a state arithmetic:** In this case, it involves observing, recording and computing the amount of resources, financial, human and material, available to a government for the purpose of governance or war. Every government needs accurate statistics to make governance easier.

**(ii) Statistics can be regarded as pieces of information:** Statistics imply data or pieces of information e.g. the age of Bayo, the height of Ike, the weight of Audu, the number of students in Mr. Bassey's class, the number of classes in JSS. 1, Federal Government College, Okigwe. Others are: number of accidents on road A for a year, number of candidates employed by company B in 1999, the number of workers retrenched during the reform programme.

**(iii) Statistics as summaries of information:** In this case, it can be used as summaries of information about a small group of individuals selected from large group for the purpose of investigating the large group. This is called sample statistics. This can be in the form of sample size, mean, median, variance, standard deviation, mode, etc. Each of these is regarded as a statistic.

**(iv) Statistics as Mathematical function or models:** In this case, it is used for comparison of two or more samples. In other words, it can be used for pair wise differences, ratios of 2-test, 2-score, t-score, t-test, f-test etc are examples.

**(v) Statistics as academic discipline:** In this case, it is regarded as a subject or field of study, in which case, it is an aspect of applied mathematics.

According to Spiegel (1972), statistics is concerned with scientific methods for collecting, organizing, summarizing, presenting and analysing data as well as drawing valid conclusions and making reasonable decisions on the basis of such analysis. statistics is further seen as the art and science of collecting, analyzing, presenting, and interpreting data. Particularly in business and economics, the information provided by collecting, analyzing, presenting, and interpreting data gives managers and decision makers a better understanding of the business and economic environment and thus enables them to make more informed and better decisions.

You can get so many definitions of statistics from so many textbooks. Since this course is not purely on statistics, we shall look at statistics as the science of decision making in the face of uncertainties. Look at Hays (1973). He says that statistics serves in two capacities:

(1) It gives methods for organizing, summarizing and communicating data, and
(2) It provides methods for making inference beyond the observations.

In summary, statistics involves observation, collection of data, organization of data, presentation of data, analysis of data, interpretation of data and decision making. You may wish to note that statistics, when used as a subject, is not the plural of statistic. A statistic is a measure which we obtain by observing the characteristics of the sample. You have learnt that we study a sample in order to make inferences about the population.

Therefore, the characteristic of the population which we estimate from a sample characteristic or statistic is called a parameter. The mean of a sample is 50. The mode of the

distribution is 45. It means that 50 is a statistic, 45 is also a statistic. You can give other examples.

## 3.2 Types of Statistics

You may have heard about different types of statistics, such as Correlation, probability, parametric, non-parametric, etc. statistics. All these have been grouped into two major types. These are descriptive and inferential statistics. In this section, you will read a brief presentation of these major types.

### 3.2.1 Descriptive Statistics

This can be described as a type of statistical application which is concerned with the organization and presentation of data in a convenient; usable and communicable form. Spiegel (1972) described it "as the set of methods serving the functions of organizing, summarizing and communicating data.

You can use descriptive statistical methods when you are interested in merely describing the characteristics of the group or the sample of study. It means that the descriptive analysis which you make will not generalize beyond the particular group or sample observed. In the same way, conclusions drawn from the study are limited and apply only to that group of study.

### 3.2.2 Inferential Statistics

These are statistical methods used for arriving at conclusions extending beyond immediate data. They are the phases of statistics which can be used to deal with conditions under which conclusions are drawn about a larger group based on data collected from some smaller group or groups chosen from and related to the larger group.

Inferential statistics can be described as a statistical procedure which makes use of sample statistics to make inferences about the population parameters. It involves the process of sampling that is representative of the population. It makes use of the aspect of inferential statistics called parametric statistics which are powerful tests that make use of the normal probability model, or making comparison involving the setting up of confidence limit, setting up of the degree of freedom etc. We shall discuss this later.

## 3.3 Benefits of the Study of Statistics

When you study statistics, you stand to derive some general benefits. These benefits focus on the useful knowledge, skills, capabilities or dispositions which you will acquire from the study of, or training in statistics. They vary, according to the extent and level of study, or training in the subject. Some of these benefits include that the study of statistics will enable you to:

1. Acquire' knowledge and skills in observation, collection, organization, communication, analysis of data, drawing inferences from the analysis of data and making sound decisions;
2. Make meaningful contributions to local, national or international debates on topical issues;
3. Read, understand and interpret communicated data, follow inferences drawn therefrom and appreciate decisions made consequent upon the inferences drawn;
4. Successfully execute empirical research. No reasonable or worthwhile empirical research can be carried out or reported without statistics for answering research questions, testing hypotheses or taking decisions and making predictions;
5. Read, interpret and make use of research reports or articles;

6. Follow and critique contributions to debates presented with facts and figures;
7. Acquire the skills and techniques for estimating, predicting and projecting into the future based on the previous and present data;
8. Draw sound conclusions based on some pieces of information that are probable or not quite certain.

**Self-Assessment Exercise**

i. What is statistics?
ii. What are the two types of statistics?

## 3.4 Organization of Data

Data collected in education can be from various sources and can be in various forms, such as: opinions, scores/marks, frequencies, verbal etc.

The data can be organized or arranged to make them meaningful. In this section, we shall look at sequencing, tables, frequency distribution tables, bar charts, etc.

### 3.4.1 Sequencing

This involves arranging the data in order of magnitude - ascending or descending order. See example below:

**Example 1:**

Given that the test scores of 10 students in statistics are:
8, 9, 2, 5, 7, 6, 4, 9, 8, 3.
This could be arranged in ascending order thus:
2, 3, 4, 5, 6, 7, 8, 8, 9, 9 or in descending order thus; 9, 9, 8, 8, 7, 6, 5, 4, 3, 2.

If the data consists of names, they can be arranged in alphabetical order. If they consists of objects, events, animals, etc. they can be arranged according to kinds, species, groups etc.

### 3.4.2 Tables

A table can be regarded as a two-dimensional representative of statistical information or data. Tables can be simple or complex as shown in the examples on the enrolment of pupils in central school Umuihi from 2000 to 2007, and Distribution of Mathematics teachers in Okigwe Zone in the year 2006.

**Example 1**

**Pupils' Enrolment in Central School, Umuihi, 2000 - 2007.**

| S/N | Year | Boys | Girls | Total |
|---|---|---|---|---|
| 1. | 2000 | . 200 | 170 | 370 |
| 2. | 2001 | 210 | 165 | 375 |
| 3. | 2002 | 230 | 170 | 400 |
| 4. | 2003 | 220 | 175 | 395 |

| | | | | |
|---|---|---|---|---|
| 5. | 2004 | 240 | 180 | 420 |
| 6. | 2005 | 225 | 170 | 395 |
| 7. | 2006 | 242 | 182 | 424 |
| 8. | 2007 | 250 | 200 | 450 |

**Example 2**

**Distribution of Mathematics Teachers in Okigwe Education Zone**

| S/N | Local Government | No, of Teachers |
|---|---|---|
| 1. | Ehime Mbano | 525 |
| 2. | Ihitte / Uboma | 425 |
| 3. | Isiala Mbano | 600 |
| 4. | Obowo – Etiti | 400 |
| 5. | Onuimo | 325 |
| 6. | Okigwe | 425 |
| | **Total** | **2,700** |

### 3.4.3 Frequency Distribution Table

A frequency distribution table shows the number of times each score, value or item occurs in a distribution. Ft consists of two columns - one for the scores/items and the other for the frequency.

**Example 3:**

The scores of some students in a Mathematics test are given below. Present the scores in a frequency table.

10, 15, 18, 12, 14, 15,20, 15, 16, 11, 12, 14, 19,20, 17, 18, 15, 13, 11, 12, 19, 13, 10, 14, 17, 19, 16,15,15,15.

**Table 18.3: Frequency Distribution Table**

| S/N | Score | Tally | Frequency |
|---|---|---|---|
| 1. | 10 | \\ | 2 |
| 2. | 11 | \\ | 2 |
| 3. | 12 | \\\ | 3 |
| 4. | 13 | \\ | 2 |
| 5. | 14 | \\\ | 3 |
| 6. | 15 | ~~\\\\~~ \\ | 7 |
| 7. | 16 | \\ | 2 |
| 8 | 17 | \\ | 2 |
| 9. | 18 | \\ | 2 |
| 10. | 19 | \\\ | 3 |
| 11. | 20 | \\ | 2 |
| | | | **30** |

Note that when you tally, each number tallied is neatly cancelled to avoid confusion.

### 3.4.4. Grouped Frequency Distribution

Some of the times, the number of scores may be so large that it becomes necessary to group several scores together. A group of score values form a class interval.

**Example 4:**

Present the scores below in a grouped frequency table.

55, 62 60, 50, 52, 58, 55, 60, 51, 55, 68, 55, 47, 39, 58, 42, 47, 42, 48, 55, 48, 46, 55, 51, 58, 65, 52, 35, 54, 55, 52, 56, 46, 65, 53, 34, 48, 50, 39, 59, 53, 52, 33, 48, 65, 60, 36, 68, 45, 62, 59, 60, 33,40,61,38.

In order to determine the interval or class size:

(i) Find the range. This is given by the highest score minus the lowest score. From the scores, we have 60 - 33 = 27.

(ii) Determine the number of groups. It has to be between 10 and 20.

(iii) Divide the range by the number e.g. 27 ÷ 12 = 3 (approximate)

(iv) Draw a table and tally the scores according to groups.

**Table 18.4: Grouped Frequency Distribution Table**

| S/N | Class Interval | Tally | Frequency | | |
|---|---|---|---|---|---|
| 1. | 66-68 | \\ | 2 | | |
| 2. | 63-65 | \\\ | 3 | | |
| 3. | 60-62 | ~~\\\\~~ \\ | 7 | | |
| 4. | 57-59 | ~~\\\\~~ | 5 | | |
| 5. | 54-56 | ~~\\\\~~ \\\\ | 9 | | 55, 62 60, 50, 52, 58, 55, 60, 51, 55, 68, |
| 6. | 51-53 | ~~\\\\~~ \\\ | 8 | | 55, 47, 39, 58, 42, 47, 42, |
| 7. | 48-50 | ~~\\\\~~ \ | 6 | | 48, 55, 48, 46, 55, 51, 58, 65, 52, |
| 8. | 45-47 | ~~\\\\~~ | 5 | | 35, 54, 55, 52, 56, 46, 65, 53, 34, |
| 9. | 42-44 | \\ | 2 | | 48, 50, 39, 59, 53, 52, 33, 48, 65, 60, |
| 10. | 39-41 | \\\ | 3 | | 36, 68, 45, 62, 59, 60, 33, 40, 61,38. |
| 11. | 36-38 | \\ | 2 | | |
| 12. | 33-35 | \\\\ | 4 | | |

# UNIT 2: METHODS OF REPRESENTING DATA AND MEASURES OF CENTRAL TENDENCY



## 1.0 INTRODUCTION

In last unit, you were exposed to the concept of statistics and organization of data. You also read through the bar chart which is a graphical way of representing data. In this unit, you will continue to be exposed to other ways of representing data. These include pie chart, histogram, frequency polygon and ogive. We will also look at the measures of central tendency. As this is not a complete course on statistics, we may not be so detailed in the presentations.

## 2.0 OBJECTIVES

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

(i) construct a pie chart using given data;
(ii) construct and describe histogram;
(Iii) draw a composite table and construct a frequency polygon;
(iv) draw a composite table and construct an ogive;
(v) calculate the mean, median and mode of a given data.

## 3.0 MAIN CONTENT

### 3.1 Pie Chart

This is used to represent both discrete and continuous data. It involves using a circle to represent a set or groups of items or scores. Each group or item is represented by a sector of the circle. The angle subtended at the centre by the sector is proportional to the frequency of the items o scores represented. It implies that the total frequencies of the set are represented by 360°.

**Example 1:**

Construct a pie chart to represent the data below:
The distribution by local government area of Basic Technology teachers in Okigwe Zone is a follows:
EHIME = 60, IHITTE/UBOM = 50, ISIALA = 65, ONUIMO = 40, OBOWO = 35, OKIGWE = 30.

To construct the pie chart:

i. Find the angle that is subtended at the centre by each group:

(a) EHIME = $\frac{60}{280} \times \frac{360}{1}$ = $77.14^0$

(b) IHITTE/UBOMA = $\frac{50}{280} \times \frac{360}{1}$ = $64.29^0$

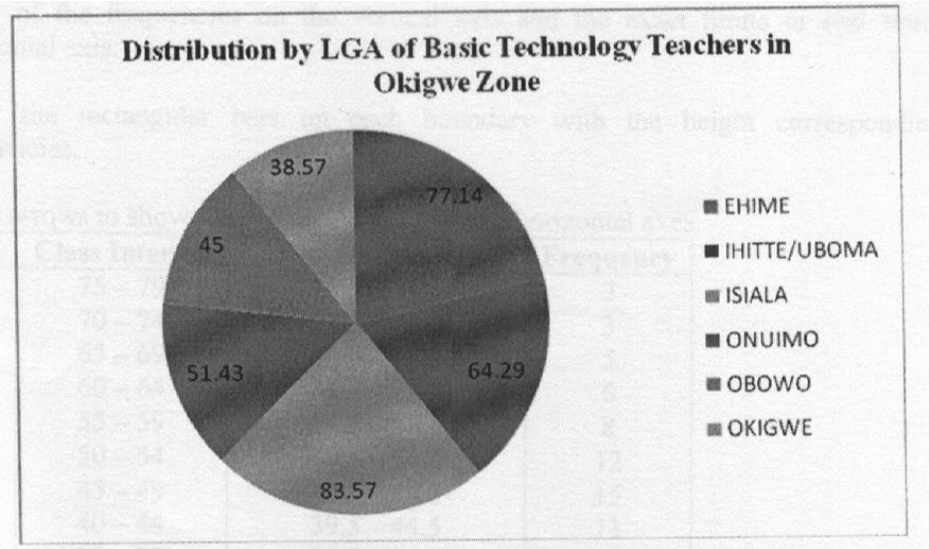
(c) ISIALA = $\frac{65}{280} \times \frac{360}{1}$ = $83.57^0$

(d) ONUIMO = $\frac{40}{280} \times \frac{360}{1}$ = $51.43^0$

(e) OBOWO = $\frac{35}{280} \times \frac{360}{1}$ = $45.00^0$

(f) OKIGWE = $\frac{30}{} \times 360$ = $38.57^0$

$\frac{}{280}$ $\frac{}{1}$




ii. With the aid of a pair of compasses, any convenient radius, draw a circle.

iii. Using your protractor, mark out the angles corresponding to each group or category o items, to the nearest degree.

i. Label the sectors of the circle corresponding to the items.

## 3.2 Histogram

In the last unit, you studied the bar chart, which is used mainly for the representation of discrete data. In the construction, you noticed that the rectangles do not touch each other. The histogram is used to represent data on a frequency distribution table like the bar chart. It is made up of rectangular bars of equal joined to one another, and it is used for continuous data. At the vertical axis, we have the frequencies and at the horizontal, we have the corresponding class intervals. The difference between the two is that, for bar chart the class intervals are used while for histogram the exact class boundaries arc used. There are two exact class boundaries—upper and lower exact class boundaries. These are obtained by subtracting 0.5 from the upper boundary and adding 0.5 to the lower boundary. Alternatively, for the exact lower limit of the first group (20 - 24), we have $\frac{19 + 20}{2} = 19.5$

And for the exact upper limit = $\frac{24 + 25}{2} = 24.5$

**Example 2**

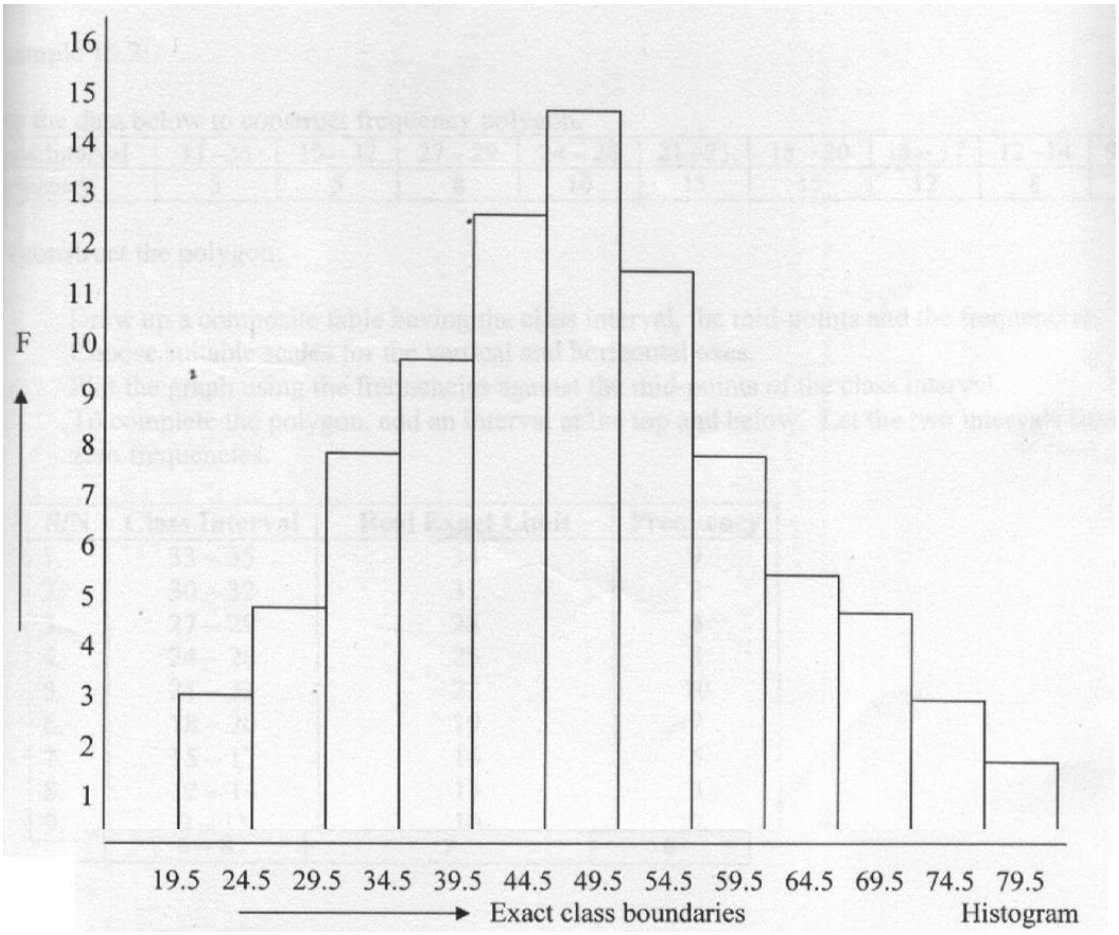
Using the data below, construct a histogram:

| Class Interval | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | 60-64 | 65-69 | 70- 74 | 75 - 79 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 3 | 5 | 8 | 10 | 13 | 15 | 12 | 8 | 6 | 5 | 3 | 2 |

To construct a histogram:

1. Compose a composite table having the class interval, the exact class limits, and the frequencies.
2. Choose suitable scales and draw the vertical and horizontal axes.
3. Mark of the frequencies on the vertical axis and the exact limits or real limits on the horizontal axis.
4. Draw the rectangular bars on each boundary with the height corresponding to the frequencies.
5. Draw arrows to show what is on the vertical and horizontal axes.

| S/N | Class Interval | Real Exact Limit | Frequency |
|---|---|---|---|
| 1. | 75-79 | 74.5 - 79.5 | 2 |
| 2. | 70-74 | 69.5 - 74.5 | 3 |
| 3. | 65-69 | 64.5 - 69.5 | 5 |
| 4. | 60-64 | 59.5-64.5 | 6 |
| 5. | 55-59 | 54.5-59.5 | 8 |
| 6. | 50-54 | 49.5 - 54.5 | 12 |
| 7. | 45-49 | 44.5-49.5 | 15 |
| 8. | 40-44 | 39.5-44.5 | 13 |
| 9. | 35-39 | 34.5-39.5 | 10 |
| 10. | 30-34 | 29.5 - 34.5 | 8 |
| 11. | 25-29 | 24.5 - 29.5 | 5 |
| 12. | 20-24 | 19.5-24.5 | 3 |
| | | | **90** |




## Self Assessment Exercise:

An Education student of NOUN spent a total of ₦30,000.00 as follows:

| | | |
|---|---|---|
| Registration | = | ₦5,000.00 |
| Course materials | = | ₦10,000.00 |
| Examinations | = | ₦5,000.00 |
| Transportation | = | ₦3,000.00 |
| Stationeries | = | ₦2,000.00 |
| Diskettes and CDs | = | ₦1, 000.00 |
| Note books | = | ₦2,500.00 |
| Typing of assignments | = | ₦l,500.00 |

Represent these expenses in a pie chart.

### 3.3 Frequency Polygon

This is a line graph plotted using the frequencies against the mid-points of the class intervals.

**Example 2**

Use the data below to construct frequency polygon.

| Class Interval | 33-35 | 30-32 | 27-29 | 24-26 | 21-23 | 18-20 | 15-17 | 12-14 | 9-11 |
|---|---|---|---|---|---|---|---|---|---|
| Frequency | 3 | 5 | 8 | 10 | 13 | 15 | 12 | 8 | 6 |

To construct the polygon:

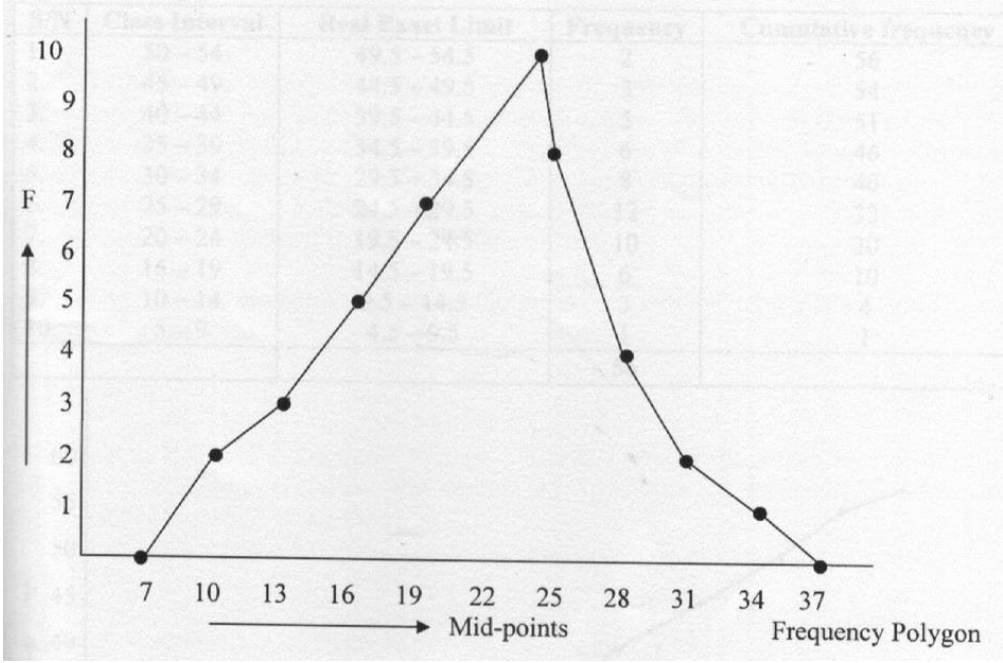
(i) Draw up a composite table having the class interval, the mid-points and the frequencies.

(ii) Choose suitable scales for the vertical and horizontal axes.

(iii) Plot the graph using the frequencies against the mid-points of the class interval.

(iv) To complete the polygon, add an interval at the top and below. Let the two intervals have zero frequencies.

| S/N | Class Interval | Real Exact Limit | Frequency |
|---|---|---|---|
| 1. | 33-35 | 34 | 9 |
| 2. | 30-32 | 31 | 2 |
| 3. | 27-29 | 28 | 4 |
| 4. | 24-26 | 25 | 8 |
| 5. | 21-23 | 22 | 10 |
| 6. | 18-20 | 19 | 7 |
| 7. | 15-17 | 16 | 5 |
| 8. | 12-14 | 13 | 3 |
| 9. | 9-11 | 10 | 2 |
| | **6-8** | **7** | **0** |




### 3.4 Ogive

This is a graph which involves the use of a smooth curve to join the Cartesian coordinate plots of cumulative frequencies against the real class boundaries. In other words, instead of the frequencies, it makes use of the cumulative frequencies. The graph gives shape like shallow 'S'.

**Example 4:**

Using the score groups below, draw an ogive or cumulative frequency curve:

| Class Interval | 5-9 | 10-14 | 15-19 | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 |
|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 1 | 3 | 6 | 10 | 12 | 8 | 6 | 5 | 8 | 6 |

To draw the ogive:

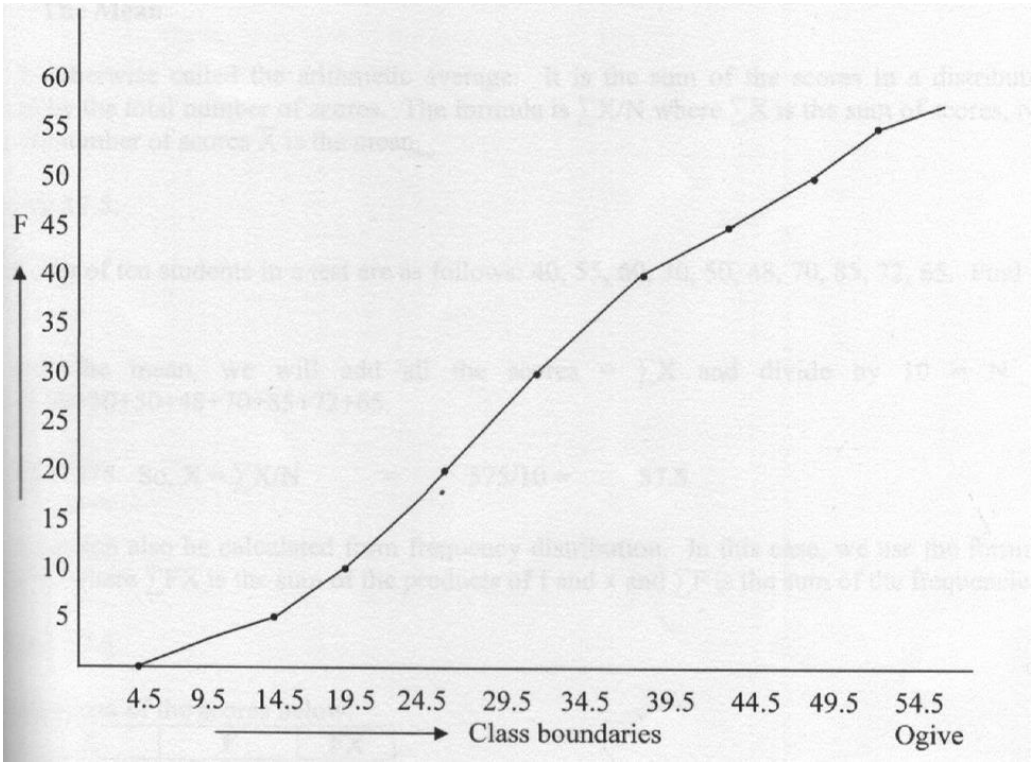
i. Compose a composite table having the class boundaries, the exact class limits, frequencies and cumulative frequencies.

ii. Choose a suitable scale to accommodate the highest cumulative frequency on the vertical axis and the class boundaries on the horizontal axis.

iii. Plot the points on the cumulative frequencies against the corresponding class boundaries.

iv. Join with a smooth curve.

| S/N | Class Interval | Real Exact Limit | Frequency | Cumulative frequency |
|---|---|---|---|---|
| I. | 50-54 | 49.5-54.5 | 2 | 56 |
| 2. | 45-49 | 44,5-49.5 | 3 | 54 |
| 3. | 40-44 | 39.5-44.5 | 5 | 51 |
| 4. | 35-39 | 34.5-39.5 | 6 | 46 |
| 5. | 30-34 | 29.5 - 34.5 | 8 | 40 |
| 6. | 25-29 | 24.5-29.5 | 12 | 32 |
| 7. | 20-24 | 19.5-24.5 | 10 | 20 |
| 8. | 15-19 | 14.5-19.5 | 6 | 10 |
| 9. | 10-14 | 9.5-14.5 | 3 | 4 |
| 10. | 5-9 | 4.5 - 9.5 | 1 | 1 |
| | | | **56** | |



## IN-TEXT QUESTIONS

Using the data below;

(i) Construct a frequency polygon, and

(ii) Construct an ogive.

| Class Interval | 10-12 | 13-15 | 16-18 | 19-21 | 22-24 | 25-27 | 28-30 | 31-33 | 34-36 | 37-39 | 40-42 | 43-45 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequencies | 2 | 4 | 6 | 10 | 7 | 12 | 8 | 5 | 0 | 4 | 3 | 1 |

### 3.5 Measures of Central Tendency

In the last sections, you studied the graphical method of representing data. The measures of central tendency provide convenient way of summarizing data. This method involves finding a single measure which is typical of a set of scores. This measure of value can be used to 'capture' or represent a whole set of scores in such a way that it becomes the representative score of the whole distribution of scores. As a teacher, you will need to be using it very often in describing the performance of your students in tests and examinations.

In statistics, the three most common of all the measures available for use are mean, median and mode. Let us discuss them in that order.

#### 3.5.1 The Mean

This is otherwise called the arithmetic average. It is the sum of the scores in a distribution divided by the total number of scores. The formula is $\Sigma X/N$ where $\Sigma X$ is the sum of scores, N is total number of scores $\bar{X}$ is the mean.

**Example 5:**

The scores often students in a test are as follows: 40, 55, 60, 30, 50, 48, 70, 85, 72, 65. Find the mean.

To find the mean, we will add all the scores = $\Sigma X$ and divide by 10 = N i.e. 40 + 55+60+30+50+48+70+85+72+65.

$\therefore \Sigma X = 375$. So $\bar{X} = \Sigma X/N = 575/10 = 57.5$

The mean can also be calculated from frequency distribution. In this case, we use the formula: ΣFX/ΣF, where ΣFX is the sum of the products of f and x and £F is the sum of the frequencies.

**Example 6:**

Find the mean of the scores below:

| **S/N** | **X** | **F** | **FX** |
|---|---|---|---|
| 1. | 30 | 2 | 60 |
| 2. | 20 | 4 | 80 |
| 3. | 15 | 4 | 60 |
| 4. | 25 | 3 | 75 |
| 5. | 10 | 8 | 85 |
| 6. | 8 | 2 | 16 |
| 7. | 5 | 6 | 30 |
| 8. | 21 | 2 | 42 |
| 9. | 12 | 1 | 12 |
| 10. | 24 | 5 | 120 |
| | | **37** | **575** |

(i) Complete the table by finding the corresponding FX i.e. FXX;

(ii) Add up F to find ΣF;

(iii) Add up FX to get ΣFX;

(iv) Divide ΣFX by ΣF - ΣFX/ΣF = 575/37 - 15.5

The mean can also be calculated when grouped frequency distribution is given.

**Example 7:**

Use the data given below to calculate the mean:

| Class Interval | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | 60-64 |
|---|---|---|---|---|---|---|---|---|
| Frequency | 2 | 5 | 6 | 7 | 10 | 6 | 3 | 2 |

(i) Complete the table by getting the mid-points X, and FX;

(ii) Use the same formula $\bar{X} = \Sigma FX/\Sigma F$.

| S/N | Class Interval | Mid-point (X) | F | FX |
|---|---|---|---|---|
| 1. | 60-64 | 62 | 2 | 124 |
| 2. | 55-59 | 57 | 3 | 171 |
| 3. | 50-54 | 52 | 6 | 312 |
| 4. | 45-49 | 47 | 10 | 470 |
| 5. | 40-44 | 42 | 7 | 294 |
| 6. | 35-39 | 37 | 6 | 222 |
| 7. | 30-34 | 32 | 5 | 160 |
| 8. | 25-29 | 27 | 2 | 54 |
| | | | 47 | 1807 |

$\Rightarrow \overline{X} = 1807/41 = 44.07$

You have seen that the mean can be calculated from both grouped and ungrouped data, using different methods. One of these methods is called the assumed mean method. It is called the short-cut.

**Example 8:**

Find the mean using the data on e.g. 7.

| S/N | Class Interval | Mid-point (X) | F | $X^1$ | $FX^1$ |
|---|---|---|---|---|---|
| 1 | 60-64 | 62 | 2 | 4 | 8 |
| 2 | 55-59 | 57 | 3 | 3 | 9 |
| 3 | 50-54 | 52 | 6 | 2 | 12 |
| 4 | 45-49 | 47 | 10 | I | 10 |
| 5 | 40-44 | 42 | 7 | 0 | 0 |
| 6 | 35-39 | 37 | 6 | -1 | -6 |
| 7 | 30-34 | 32 | 5 | -2 | -10 |
| 8 | 25-29 | 27 | 2 | -3 | -6 |
| | | | 47 | | 17 |

(i) Take away group mark as the assumed mean, and code it 0 as shown in column $X^1$.

(ii) Code every other mark above from 1, 2, 3 etc. and below -1, -2, etc.

(iii) Find the $FX^1$ and sum up.

(iv) Use the formula $AM + (\Sigma FX^1/\Sigma F)^i = \overline{X}$.

$\Rightarrow \overline{X} = 42 + (17/41) = 42 + + 2.073 = 44.073$

$= 44.07$

### 3.5.2 The Median

This is the score in the distribution above and below which 50% of the scores lie. It is the middle score which divides the set of scores into two equal halves. In order to get the median, the scores must be arranged in an ordering -ascending or descending.

**Example 9:**

Find the median of the sets of scores:

(a) 9,7, 15,10, 11,8,2,4,3.

(b) 5, 9, 8, 7, 3, 2, 4, 6, 5, 8.

In example (a), simply arrange in ascending order. By this, we have: 2, 4, 5, 7, 8, 9, 10, 11, 15. By counting, the middle number, which is 8 is the median.

In example (b), you will notice that the number is even. You will therefore arrange in order, by counting, the two middle numbers are taken, added and divided by two.

We have:

2, 3, 4, 5, 5, 6, 7, 8, 8, 9.

The median is $\frac{5+6}{2} = \frac{11}{2} = 5.5$

When grouped data are given, the median is calculated using the formula

$$\ddot{X} = L + \frac{(N\,|2 - cfb)^i}{fw}$$

where L is the lower boundary of the median class;

N is the number of scores;

cfb is the cumulative frequency below the median class;

fw is the frequency within the median class.

**Example 10:**

Use the data below to find the median:

| S/N | Class Interval | F | FX |
|---|---|---|---|
| 1. | 85-89 | 1 | 52 |
| 2. | 80-84 | 2 | 51 |
| 3. | 75-79 | 3 | 49 |
| 4. | 70-74 | 5 | 46 |
| 5. | 65-69 | 7 | 41 |
| 6. | 60-64 | 8 | 34 |
| 7. | 55-59 | 10 | 26 |
| 8. | 50-54 | 6 | 16 |
| 9. | 45-49 | 5 | 10 |
| 10. | 40-44 | 4 | 5 |
| 11. | 35-39 | 0 | 1 |
| 12. | 30-34 | 1 | 1 |
| | | **52** | |

(i) N/2 = 52/2 = 26

(ii) Find the class where 26 lies in the cummulative frequency i.e. 55 - 59. This is the median class.

(iii) $\ddot{X} = L + \frac{(N\,|2 - cfb)^{i}}{fw} = 54.5 + \frac{(26 - 10)^{5}}{10}$

$(10)^{5}$

$$= 54.5 + \frac{\quad}{10} = 54.5 + 5$$

$$= 59.5$$

### 3.5.3 The Mode

This is the most frequently occurring score or scores in a distribution, It is the most frequent score which can be easily determined by inspection. But in some distributions, you may have two modes. This is called bimodal; any distribution with more than two modes is called multi-modal.

Now, let us look at how to find the modes in the examples below:

**Example 11:**

Find the mode in the distribution below:

20, 30, 21, 45, 30, 25, 33, 35, 30, 22, 29, 30.

By inspection, you will see that 30 appeared 4 times. It is the mode because no other score appeared up to that.

**Example 12:**

Find the mode in the frequency table given below:

| X | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|
| F | 1 | 2 | 4 | 5 | 8 | 6 | 4 | 3 | 1 | 1 |

Again, by inspection, you will see that the highest occurring frequency in the above distribution is 8, and the value is 6. Therefore, 6 is the mode.

For a grouped data, the mode is calculated using the formula below:

$$\dot{\ddot{X}} = L + \frac{(d^1)^i}{d^1 + d^2}$$

where L is the exact lower limit of the modal class;

$d^1$ is frequency of the modal class minus frequency of the class preceding or before the modal class;

$d^2$ is frequency of the modal class minus frequency of the class immediately after the modal class.

**Example 13:**

Find the mode in the frequency table given below:

| S/N | Class Interval | F |
|---|---|---|
| 1. | 85-89 | 3 |
| 2. | 80-84 | 3 |
| 3. | 75-79 | 8 |
| 4. | 70-74 | 10 |
| 5. | 65-69 | 12 |
| 6. | 60-64 | 7 |
| 7. | 55-59 | 5 |
| 8. | 50-54 | 2 |

(i) Locate the modal class i.e., 65 - 69.

(ii) Using the formula $L + \frac{(\frac{1}{2}i}{d^1 + d^2} = 2$

where L = 64.5, i = 5, $d^1$ = 12 - 7 - 5, $d^2$ = 12 - 10,

(iii) $X = 64.5 + \frac{(5)}{5}$

$\frac{(5)}{2}$

$5 + 2$

$64.5 + \frac{}{7}$

= 64.5 + 3.571 = 68.07

## IN-TEXT QUESTIONS

i. Define mean, median and mode.

ii. Find the mean, median and mode of the distribution given below:

10, 7, 8, 9, 6, 9, 3, 2, 9, 5, 1.

## 4.0 CONCLUSION

You have noticed that data by themselves convey little or no meaning until they are summarised and described. Some methods of representing data have been presented and the measures of central tendency, which form the central reference value that is usually close to the point of greatest concentration of the measurement, and which may in some sense be thought of typify the whole set, have also been presented. In the next unit, we shall look at other statistical measures.

## 5.0 SUMMARY

In this unit, you have been able to go through the other methods of representing data which you started in unit sixteen of this module. You have seen that the pie chart uses a circle to represent a set of data or groups of items. In other words, it can be used for both discrete and continuous data. You also went through the histogram, which is made up of rectangular bars of equal width joined to one another, It is used for continuous data. The frequency polygon is a line graph plotted using the frequencies against the mid-points of the class intervals. The ogive uses the cumulative frequencies against the exact "class boundaries. We have two types of ogives - 'less than' ogive and 'greater than' ogive.

You have equally worked through the measures of central tendency. The three measures are the mean, the median and the mode. You have seen how to calculate these measures. In the next unit, we shall look at other measures.

## 6.0 TUTOR MARKED ASSIGNMENT

Find the mean, the median and the mode of the data given below;

| Class Interval | 15- 19 | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | 60-64 | 65-69 | 70-74 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequencies | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 4 | 3 | 2 | 1 | 1 |

**REFERENCES**

Anaekwe, M. C. (2016), Basic Research Methods and Statistics in Education and Social Sciences Enugu: Podiks Printing and Publishing Company.

Dey, I. (2016). Quantitative data analysis. London: Taylor and francis Group.

Ogomaka, P. M. C. (2014) Descriptive Statistics for Research Students. Owerri: Peacewise

## UNIT 3: MEASURE OF VARIABILITY OR SPREAD



### 1.0 INTRODUCTION

In the last unit, you worked through the measures of central tendency. In addition to those measures, researchers are also interested to know how the scores are spread or scattered in the distribution.

So the measures of variability indicate the degree to which a set of scores differs from each other in the distribution. These measures present a measure of homogeneity within the group of scores.

In this unit, we shall look at the range, the quartiles, the percentiles, the variance and the standard deviation.

### 2.0 OBJECTIVES

After working through this unit, you will be able to:

(i) find the range in a given set of scores;

(ii) explain and find the quartiles in a distribution;

(iii) find the percentiles in a given set of scores;

(iv) calculate the variance in a given set of scores;

(v) calculate the standard deviation in a distribution.

### 3.0 MAIN CONTENT

### 3.1 The Range

This is the simplest and crudest measure of variability which measures the distance between the highest and the lowest scores in a distribution of scores. It is calculated by subtracting the lowest score from the highest score in the distribution of scores, plus one.

**Example 1:**

Find the range of the scores below:

30, 45, 20, 32, 70, 85, 90, 44, 60.

You will notice that the lowest score is 20 and the highest score is 90. So, $X_h - X_L + 1 = 90 - 20 + 1 = 71$. The range is 71.

You would have seen that the range is affected by the two extreme scores. Therefore, it is an unstable and unreliable method of determining the spread of scores. Because of this limitation, it is seldomly used as an indicator of the spread.

## 3.2 The Quartiles

These are score points or values which subdivide a given distribution into four equal parts. In other words, the number of scores in anyone of the four groups is equal to the number of scores in any other of the remaining three groups.

There are only three quartiles for any given 'distribution. These are the first quartile $Q_1$, second quartile $Q_2$ and third quartile $Q_3$. This can be illustrated below;

### 3.2.1 Calculation of the Quartiles

The quartiles can be calculated in a grouped data using the formula

$$Qi = L + \frac{\left[i(N\ |4) - cfb\right]^c}{fw}$$

Where i = 1, 2, 3, (i.e. the quartiles)

N = Σf = sample size

L = lower class boundary of the quartile class

cfb = cummulative frequency below the quartile class

fw = frequency of the quartile class

c = class interval size.

**Example:**

Find $Q_1$ and $Q_3$ in the distribution below:

| S/N | Class Interval | F | |
|---|---|---|---|
| 1. | 50 – 54 | 1 | 34 |
| 2. | 45 – 49 | 2 | 33 |
| 3. | 40 – 44 | 2 | 31 |
| 4. | 35 – 39 | 5 | ← 29 |
| 5. | 30 – 34 | 8 | 24 |
| 6. | 25 – 29 | 6 | 16 |
| 7. | 20 – 24 | 4 | ← 10 |
| 8. | 15 – 19 | 3 | 6 |
| 9. | 10 – 14 | 2 | 3 |
| 10. | 5 – 9 | 1 | 1 |
| | | 34 | |

Step: (i) Find the cumulative frequencies (CF)

(ii) Divide 34 by 4 = $\frac{34}{4} = 8.5$

(iii) Apply the formula $Q_1 = L + \frac{(i(\frac{N}{4}) - cfb)^c}{fw}$

For $Q_1$: 8.5 lies in the class 20 – 24

$\therefore$ 20 – 24 is the quartile class.

So, L – 19.5, fw = 4, cfb = 6

Then, $Q_1 = L + \frac{(\frac{N}{4}) - cfb)^c}{fw} = 19.5 + \frac{(8.5 - 6)^5}{4} = 19.5 + \frac{(2.5)^5}{4}$

$= 19.5 + 3.125 = 22.625$

For $Q_3$, $Q_3 = L + \frac{(\frac{N}{4}) - cfb)^c}{fw}$ = (35-39) in the class.

$$= 34.5 + \frac{(3 \times 8.5 - 24)^5}{5} = 34.5 + \frac{(1.5)^5}{5}$$

$$= 34.5 + 1.5 = 36.0$$

### 3.2.2 Interquartile Range

In the last subsection, you learnt that the quartiles divide the distribution of scores into four equal parts. The inter-quartile range describes the distance between the first quartile $Q_1$ and the third quartile $Q_3$. It shows the scores that the included in the middle 50% or half of the scores in the distribution. It is found using the formula $Q_3 - Q_1$. For instance, in the example 18.2 above $Q_3 = 36.0$ and $Q_1 = 22.625$. The inter-quartile range (IQR) is $36.0 - 22.625 = 13.775$.

### 3.2.3 Quartile Deviation or Semi-interquartile Range

The quartile deviation otherwise called semi-interquartile range is described as half the value of the interquatile range. It is calculated using the formula:

$$\frac{Q_3 - Q_1}{2}$$

For instance, in the example given above, the semi-interquartile range will be given by:

$$\frac{36.0 - 22.625}{2} = \frac{13.735}{2} = 6.87$$

## 3.3 The Percentiles

These are score points along the score line which divide a distribution of scores into hundred subgroups. The subgroups are divided in such a way that they are equal or the same. It is calculated in the same way as the quartiles, but instead of dividing N by 4, you divide by 100. Thus,

$$Pi = L + \frac{(i(\frac{N}{100}) - cfb)^c}{fw}$$

### 3.3.1 The Deciles

These are score points in a distribution which divide the distribution of scores into Ten equal parts. As in the percentile or quartile the calculation is the same. The formula is:

$$Di = L + \frac{(i(\frac{N}{10}) - cfb)^c}{fw}$$

Note that $Q_1 = P_{25}$, $Q_2 = D_5 = P_{50}$ and $Q_3 = P_{75}$

### 3.4 The Variance ($S^2$) and the Standard Deviation (S)

These two measures of variability are directly related. They are the most common, the most reliable estimate of dispersion or spread. They give relative measure to the degree to which each score differs from the mean of the distribution. The standard deviation is the square root of the variance. It is widely used than many other statistical operations.

To calculate the variance and standard deviation, the following steps are applied:

1. calculate the mean of the scores.
2. subtract the mean from each score or class midpoint (if grouped)
3. square each of the differences or deviations - $(X - \overline{X})^2$ or $d^2$ or $x^2$
4. Multiply each square deviation by the corresponding frequency, the result is f $(X - \overline{X})^2$ or $fd^2$ or $fx^2$.
5. Sum up the result in step (iv) above to obtain $\Sigma f(X - \overline{X})^2$
6. Divide the result of the sum by total number of scores N or the sum of the frequencies i.e. $\Sigma f \frac{(X - \overline{X})^2}{\sum f}$ or $\sqrt{\Sigma_f \frac{(X - \overline{X})^2}{\Sigma f}}$. This is the deviation method.

There is also the raw score method otherwise called the machine approach. We shall look at it after the deviation method. Now, let us take some examples.

| S/No | X | F | fX | $(X - \overline{X})$ | $(X - \overline{X})^2$ | $f(X - \overline{X})^2$ |
|---|---|---|---|---|---|---|
| 1 | 11 | 1 | 11 | -4.97 | 24.70 | 24.70 |
| 2 | 12 | 2 | 24 | -3.97 | 15.76 | 31.52 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 3 | 13 | 4 | 52 | -2.97 | 8.82 | 35.28 |
| 4 | 14 | 7 | 98 | -1.97 | 3.88 | 27.16 |
| 5 | 15 | 10 | 150 | -0.97 | 0.94 | 9.40 |
| 6 | 16 | 12 | 192 | 0.03 | 0.00 | 0.00 |
| 7 | 17 | 11 | 187 | 1.03 | 1.06 | 11.66 |
| 8 | 18 | 6 | 108 | 2.03 | 4.12 | 24.72 |
| 9 | 19 | 4 | 76 | 3.03 | 9.18 | 36.72 |
| 10 | 20 | 3 | 60 | 4.03 | 16.24 | 48.72 |
| | | **60** | **958** | | | **249.88** |

**Example**

Find the variance and standard deviation of the following scores:

Steps: (i) Find the mean = $\frac{\Sigma fx}{\Sigma f} = \frac{958}{60} = 15.97$

(ii) Find the deviation = $(X - X)$

(iii) Find the square deviations.

(iv) Multiply the square deviations by the frequency to obtain $f(X - X)^2$

(v) Find $\Sigma f (X - X)^2 = 249.88$

(vi) Divide by $\Sigma f$ or N to get variance.

(vii) Find the square root

$$\frac{249.88}{60} = 4.646667$$

$S^2 = 4.16$

$\therefore \sqrt{4.16} = 2.04$

You can also use the raw score approach. Let us use the raw score approach for the same set of scores in Example.

| S/N | X | f | Fx | $X^2$ | $Fx^2$ |
|---|---|---|---|---|---|
| 1 | 11 | 1 | 11 | 121 | 121 |
| 2 | 12 | 2 | 24 | 144 | 288 |
| 3 | 13 | 4 | 52 | 169 | 676 |
| 4 | 14 | 7 | 98 | 196 | 1372 |
| 5 | 15 | 10 | 150 | 225 | 2250 |
| 6 | 16 | 12 | 192 | 256 | 3072 |
| 7 | 17 | 11 | 187 | 289 | 3179 |
| 8 | 18 | 6 | 108 | 324 | 1944 |
| 9 | 19 | 4 | 76 | 361 | 1444 |
| 10 | 20 | 3 | 60 | 400 | 1200 |
| | | **60** | **958** | | **15546** |

Step:

(i) Complete the composite table as shown.

(ii) For variance ($S^2$) use the formula = $\frac{(\Sigma f)\,\Sigma fx^2 - (\Sigma fx)^2}{(\Sigma f)^2}$ or $\frac{N\Sigma fx^2\ (\Sigma fx)^2}{N^2}$

Substituting, we have: $S^2 = \frac{60 \times 15546 - 958^{\ 2}}{3600}$ or $\frac{14996}{3600}$

$= 4.165$

For standard deviation, S. Find the square root of the variance i.e. $\sqrt{4.165} = 2.04$

**IN-TEXT QUESTION**

Find the variance and standard deviation of the scores on Business Education students in BED212

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| X | 70 | 74 | 78 | 82 | 86 | 90 | 94 | 98 | 102 | 106 |
| f | 4 | 9 | 16 | 28 | 45 | 66 | 85 | 72 | 54 | 38 |

Find the variance and standard deviation of the following:

Sometimes, you may be given grouped scores. The same method is used. The only different is that you have to find and use the midpoints of the groups or class intervals as your score X.

### 4.0 Conclusion

In this unit, you have gone through the other measures which are used to determine the extent of spread or variability in a given set of scores. They represent a measure of homogeneity within a group of scores. The standard deviation is applied in most other statistical tests.

## 5.0 SUMMARY

You have seen that the range is a measure of the distance between the highest and the lowest scores in a distribution. The quartiles are score points which divide the distribution into four equal parts. We have $Q_1$, $Q_2$ and $Q_3$. The percentiles divide the distribution into hundred equal parts,. The deciles divide the distribution into ten equal parts. You have also gone through variance and standard deviation which are the most reliable estimate of dispersion or spread. The standard deviation is the square root of the variance.

In the next unit, we shall be looking at the measures of association.

### 6.0 TUTOR MARKED ASSIGNMENT

In the data below, find:

(i) The semi-interquartile range, and

(ii) The standard deviation

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 13 | 13 | 17 | 7 | 22 | 22 | 26 | 17 | 13 |
| 14 | 16 | 7 | 6 | 18 | 20 | 10 | 17 | 11 |
| 10 | 17 | 11 | 10 | 15 | 16 | 8 | 16 | 21 |

## 7.0 REFERENCES


Anaekwe, M. C. (2016) Basic Research Methods and Statistics in Education and Social Science. Enugu: Podiks Printing and Publishing Company.

David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, James J. Cochran. (2018). *Statistics for Business and Economics.* (Thirteenth Edition, Revised). United States of America: Cengage Learning.

Williams, M. (2016): Research Methods Knowledge Base. London: Taylor and francis Group

Pallant, J. (2015). *SPSS Survival Manual: A step by step guide to data analysis using SPSS.* (4th edition). China: Everbest Printing Co

Olatian, S. O and Nwoke, G. I. (2015). Practical Research Methods in Education. Onitsha: Summer Education Publishers.

# UNIT 4 MEASURES OF ASSOCIATION/CORRELATION



## 1.0 INTRODUCTION

In the previous units, we have focused on sample scores from one variable or distribution of scores from one variable.

In this unit, you will learn about matched or paired sets of scores. Specifically, measures of association show the degree of relationship between two or more variables.

We shall be looking at some of these measures or the statistics for describing the extent of correlation or 'going together' of some attributes or characteristics possessed by a sample of individuals. This degree of relationship between the attributes or variables is expressed as a coefficient of correlation. The result of this unit will teach you the most common types of correlation which are Pearson Product Moment and Spearman Rank Order.

## 2.0 OBJECTIVES

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

(i) define correlation;
(ii) illustrate the scatter-grams of various correlations;
(iii) calculate the Pearson r;
(iv) calculate the Spearman rho.

## 3.0 MAIN CONTENT

### 3.1 The Concept of Correlation

Correlation refers to the extent or degree of relationship between two variables. The index showing the degree of such relationship between the two variables is called correlation-coefficient.

The value obtained from correlation will help you as a researcher to know whether variations in one set of scores lead to variations in another set of scores. It will also help you to know the extent to which this variation takes place.

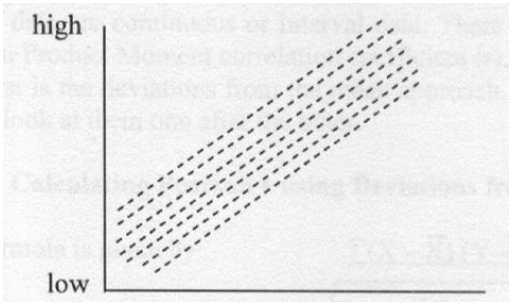
Correlation values ranges from -1 to +1. It means that a correlation coefficient of-1 indicates a perfect negative relationship, while +1 shows perfect positive relationship and 0 correlation coefficient implies no relationship at all. Many types of correlation coefficients exist. You can use any type, but this will depend on the following:

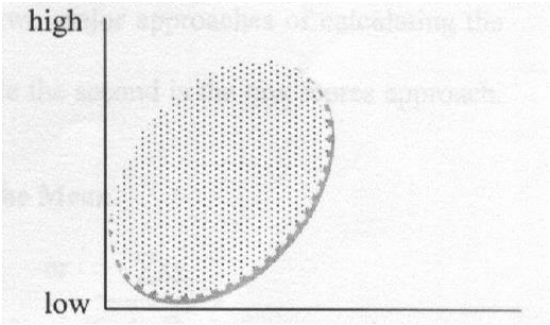
(i) the type of measurement scale in which the variables are;
(ii) the nature of the distribution (i.e. continuous or discrete);
(iii) the characteristics of the distribution scores,

### 3.2 Scatter-grams of Various Correlations

A scatter-gram is a shortened form of scatter diagram. It shows the plots on the Cartesian coordinate plane of two sets of scores of individuals of a sample with respect to two attributes which are usually denoted by X and Y.

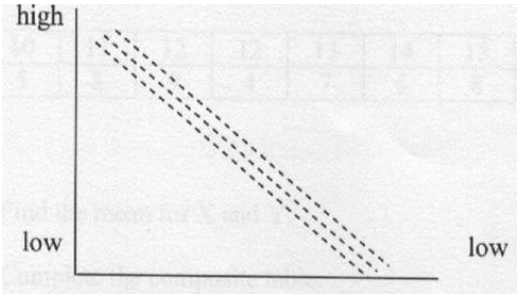
**i. Positive Relationship:** This suggests that individuals having high scores in one variable also have high scores in the other variable. It also implies that those individuals who have low scores in one variable also have low scores in the other variables.

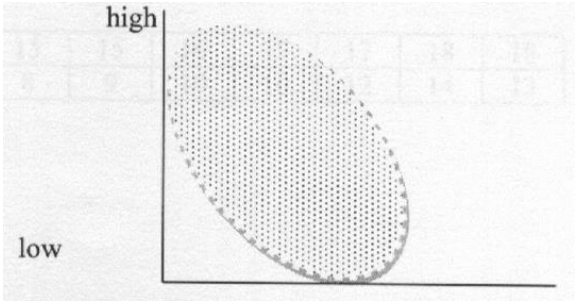



(a) r = Perfect Positive Relationship (b) r = Moderate Positive Relationship

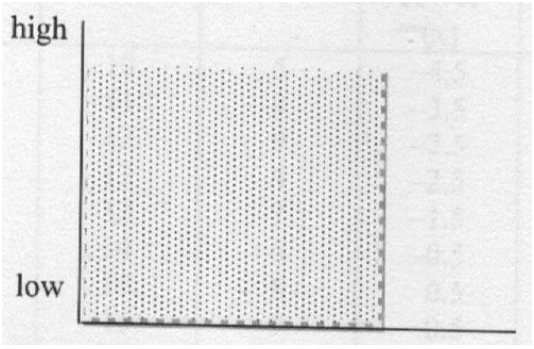
**ii. Negative Relationship:** As you can see, this is the opposite of positive relationship. It suggests that individuals scoring high on one variable score low on another variable. It also implies that those who score low on one variable score high on the other variable.




(c) r = Perfect Negative Relationship (d) r = Moderate Negative Relationship

**iii. Zero Relationship:** This suggests the absence of any relationship. There is no relationship between scores on the two variables.




(e) r = No Relationship

### 3.3 Pearson Product Moment Correlation Coefficient (r)

This type of correlation coefficient, named after the man who developed it, is used when the two sets of data are continuous or interval data. There are two major approaches of calculating the Pearson Product Moment correlation coefficient (r).
The first is the deviations from the mean approach, while the second is the raw scores approach. Let us look at them one after the other.

#### 3.3.1 Calculating Pearson r using Deviations from the Mean

The formula is given by: $$\frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2\sum(Y-\bar{Y})^2}} \text{ or } \frac{\sum xy}{\sqrt{(\sum X^2)(\sum Y^2)}} y$$

where $x = X-\bar{X}$, $y = Y-\bar{Y}$

**Example 1:**

Using the data below, calculate the Pearson r.

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X | 10 | 11 | 12 | 12 | 13 | 14 | 15 | 15 | 16 | 17 | 17 | 18 | 18 |
| Y | 5 | 8 | 9 | 4 | 7 | 6 | 8 | 9 | 10 | 10 | 12 | 14 | 13 |

Step:

(i) Find the mean for X and Y.

(ii) Complete the composite table.

(iii) If $\Sigma xy = 80.90$, $\Sigma x^2 = 87.25$, $\Sigma y^2 = 107.72$, Then $\frac{\sum xy}{\sqrt{\sum x^2 y^2}} = r$

| | X | Y | X-X | Y-Y | XY | $X^2$ | $Y^2$ |
|---|---|---|---|---|---|---|---|
| | | | (x) | (y) | | | |
| 1 | 10 | 5 | ^.5 | -3.8 | 17.10 | 20.25 | 14.44 |
| 2 | 11 | 8 | -3.5 | -0.8 | 2.80 | 12.25 | 0.64 |
| 3 | 12 | 9 | -2.5 | 0.2 | -0.50 | 6.25 | 0.04 |
| 4 | 12 | 4 | -2.5 | -^.8 | 12.00 | 6.25 | 23,04 |
| 5 | 13 | 7 | -1.5 | -1.8 | 2.70 | 2.25 | 3.24 |
| 6 | 14 | 6 | -0.5 | -2.8 | 1.40 | 0.25 | 7.84 |
| 7 | 15 | 8 | 0.5 | -0.8 | -0.40 | 0.25 | 0.64 |
| 8 | 15 | 9 | 0.5 | 0.2 | 0.10 | 0.25 | 0.04 |
| 9 | 16 | 10 | 1.5 | 1.2 | 1.80 | 2.25 | 1.44 |
| 10 | 17 | 10 | 2/5 | 1.2 | 3.00 | 6,25 | 1.44 |
| 11 | 17 | 12 | 2.5 | 3.2 | 8.00 | 6.25 | 10.24 |
| 12 | 18 | 14 | 3.5 | 5.2 | 18.20 | 12.25 | 27.04 |
| 13 | 18 | 13 | 35 | 4.2 | 14.70 | 12.25 | 17.64 |
| | **188** | **115** | | | **80.90** | **87.25** | **107.72** |
| | **14.5** | **8.8** | | | | | |

$$r = \frac{80.90}{\sqrt{87.95 \times 107.72}} = \frac{80.90}{\sqrt{9,398.57}} = \frac{80.90}{96.945}$$

r = 0.83

### 3.3.2 Calculating Pearson r using the Raw Score Method

The formula is given by $r = \frac{N\sum XY - \sum X \sum Y}{\sqrt{N\sum X^2 - \left(\sum X^2\right) - \left(N\sum Y^2 - \left(\sum Y\right)^2\right)}}$

where $x = X - \overline{X}$ , $y = Y - \overline{Y}$

**Example 2:**

Let us use the same data in example 19.1.

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X | 10 | 11 | 12 | 12 | 13 | 14 | 15 | 15 | 16 | 17 | 17 | 18 | 18 |
| Y | 5 | 8 | 9 | 4 | 7 | 6 | 8 | 9 | 10 | 10 | 12 | 14 | 13 |

Steps:

i. Complete the composite table.

ii. If N = 13, $\Sigma X = 188$, $\Sigma Y = 115$, $\Sigma XY = 1744$, $\Sigma X^2 = 2806$ and $\Sigma Y^2 = 1125$, then:

$$r = \frac{N\sum XY - \sum X \sum Y}{\sqrt{N\sum X^2 - \left(\sum X^2\right) - \left(N\sum Y^2 - \left(\sum Y\right)^2\right)}}$$

| S/N | X | Y | XY | $X^2$ | $Y^2$ |
|---|---|---|---|---|---|
| 1 | 10 | 5 | 50 | 100 | 25 |
| 2 | 11 | 8 | 88 | 121 | 64 |
| 3 | 12 | 9 | 108 | 144 | 81 |
| 4 | 12 | 4 | 48 | 144 | 16 |
| 5 | 13 | 7 | 91 | 169 | 49 |
| 6 | 14 | 6 | 84 | 196 | 36 |
| 7 | 15 | 8 | 120 | 225 | 64 |
| 8 | 15 | 9 | 135 | 225 | 81 |
| 9 | 16 | 10 | 160 | 256 | 100 |
| 10 | 17 | 10 | 170 | 289 | 100 |
| 11 | 17 | 12 | 204 | 289 | 144 |
| 12 | 18 | 14 | 252 | 324 | 196 |
| 13 | 18 | 13 | 234 | 324 | 169 |
| Σ | **188** | **115** | **1744** | **2806** | **1125** |
| | **14.5** | **8.8** | | | |

$$r = \frac{13\times1744 - 188\times115}{\sqrt{13\times 2806 - 188^2 \times 13\times1125 - 115^2}}$$

$$= \frac{22672 - 21620}{\sqrt{36478 - 35344\times14625 - 13225}}$$

$$= \frac{1052}{\sqrt{1134\times1400}}$$

$$= \frac{1052}{1260} = 0.83$$

You can see that the two approaches give the same result. This is because the formula of the raw scores method is derivable from the formula of the deviations from the mean method. You will have to note that when the scores are large and the means of X and Y are whole numbers, the deviations from the mean method becomes simper to handle. But when the means of X and Y are not whole numbers the raw score method is preferred.

**IN-TEXT QUESTION**

Use any method to calculate the Pearson r of the data:

| S/N | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| X | 51 | 44 | 70 | 32 | 65 | 67 | 19 | 71 | 45 | 80 |
| Y | 49 | 41 | 45 | 31 | 50 | 61 | 11 | 64 | 21 | 75 |

## 3.4 Spearman Rank Order Correlation Coefficient - rho

This correlation coefficient was developed independently by Spearman and Brown. This is why it is sometimes referred to as Spearman-Brown Rank Order Correlation Coefficient.

It is more popularly known as Spearman rho, because Spearman was the first to publish it. It is an approximation of the Pearson r. It is used when the scores in each variable arc ranked in the same direction, with respect to magnitude.

So, in the use of Spearman rho, ranking is emphasized. It must be done and correctly too.

### 3.4.1 Calculation of Spearman Rank Order Correlation

The formula is given by: $\text{rho} = 1 - \frac{6\sum D^2}{N(N^2 - I)}$

**Example 3:**

Calculate the rho of the data presented below:

| S/N | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| X | 51 | 44 | 70 | 32 | 65 | 67 | 19 | 71 | 45 | 80 |
| Y | 49 | 41 | 45 | 31 | 50 | 61 | 11 | 64 | 21 | 15 |

| S/N | X | 'Y | RX | RY | D | $D^2$ |
|---|---|---|---|---|---|---|
| 1 | 51 | 49 | 6 | 5 | 1 | 1 |
| 2 | 44 | 41 | 8 | 7 | 1 | 1 |
| 3 | 70 | 45 | 3 | 6 | -3 | 9 |
| 4 | 32 | 31 | 9 | 8 | 1 | 1 |
| 5 | 65 | 50 | 5 | 4 | 1 | 1 |
| 6 | 67 | 61 | 4 | 3 | 1 | 1 |
| 7 | 19 | 11 | 10 | 10 | 0 | 0 |
| 8 | 71 | 64 | 2 | 2 | 0 | 0 |
| 9 | 45 | 21 | 7 | 9 | -2 | 4 |
| 10 | 80 | 75 | 1 | 1 | 0 | 0 |
| Σ | | | | | | **10** |

Steps:

i. Complete the composite table by getting the ranks and the differences between the ranks.

ii. Apply the formula: rho = $1-\frac{6\sum D^2}{N\left(N^2-1\right)}$

rho = $1-\frac{6\times 18}{10\left(10^2-1\right)}$ = $1-\frac{108}{10\times 99}$ = $1-\frac{108}{990}$

= 1 – 0.109 = 0.891

### 3.5 Point Biserial Correlation Coefficient - rpbi

You have worked through the Pearson r and Spearman rho. Let us close this unit with the point biserial correlation coefficient which is used when one variable has dichotomized values. .Typical examples of variables which can use rpbi are scores and sex.

**Example 4:**

| S/N | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X | 10 | 15 | 11 | 13 | 12 | 18 | 20 | 14 | 16 | 17 | 09 | 07 |
| Y | G | B | G | B | G | G | B | G | B | B | B | B |

The formula for this is given by: rpbi $= \frac{\overline{X}p - \overline{X}q}{St}\sqrt{pq} = \frac{\overline{X}p - \overline{X}t}{St}\sqrt{p\,q}$

where $\overline{X}p$ = mean score of the continuous variable of the subgroup that belongs to the natural dichotomy p.

Xq = mean score of the continuous variable of the subgroup that belongs to the natural dichotomy q.

st = standard deviation of the total scores for the whole group on the continuous variable.

p = proportion of the number of members in subgroup p.

q = proportion of the number of members in subgroup q.

Now, let us look at the steps you can follow:

i. Find $\bar{X}p$ = mean for the proportion of boys and the group.

$$= \frac{15+13+20+16+17+09+07}{7} = \frac{97}{7} = 13.86$$

ii. Find $\bar{X}q$ = mean for the proportion of girls in the group.

$$= \frac{10+11+12+18+14}{5} = \frac{65}{5} = 13.0$$

iii. Find p= 7/12 = 0.58

iv. Find q = 5/12 = 0.42.

v. Find St.

| S/N | X | X-X | $(X - \bar{X})^2$ |
|---|---|---|---|
| 1 | 10 | -3.5 | 12.25 |
| 2 | 15 | 1.5 | 2:25 |
| 3 | 11 | -2.5 | 6.25 |
| 4 | 13 | -0.5 | 0.25 |
| 5 | 12 | -1.5 | 2.25 |
| 6 | 18 | 4.5 | 20.25 |
| 7 | 20 | 6.5 | 42.25 |
| 8 | 14 | 0.5 | 0.25 |
| 9 | 16 | 2.5 | 6.25 |
| 10 | 17 | 3.5 | 12.25 |
| 11 | 09 | -4.5 | 20.25 |
| 12 | 07 | -6.5 | 42.25 |
| **$\Sigma$x** | **162** | **13.5** | **167.00** |

$$St = r \sqrt{\frac{\sum(X - \bar{X})^2}{n}} = \sqrt{\frac{167}{12}} = \sqrt{13.92}$$

$$= \quad 3.73$$

$$\therefore \text{ rpbi} = \frac{\bar{X}p - \bar{X}q}{St}\sqrt{pq} = \frac{13.86 - 13.0q}{3.73}\sqrt{0.58 \times 0.42}$$

$$= \quad 0.2305563 \text{ x } 0.493558$$

$$= \quad 0.1137963 \qquad = \quad 0.11$$

**IN-TEXT QUESTION**

Find the rpbi of the following data:

| S/N | i | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X | 60 | 40 | 55 | 20 | 70 | 35 | 48 | 15 | 30 | 57 | 65 | 25 | 30 |
| Y | G | B | B | G | B | G | B | G | B | G | B | G | G |

## 4.0 CONCLUSION

A very good number of research studies tend to determine the nature and scope of relationships which exist between two or more variables being investigated. In this unit, you have seen that the degree of relationship which exists between variables is referred to as correlation. You have also noted that the statistical index of measuring the relationship is called correlation coefficient. This correlation coefficient presents a picture of how a change in one variable results in a change in the corresponding correlated variable. The result of the correlation tests can be used for predictive purposes. But they cannot be used for establishing a cause-effect relationship between two variables.

## 5.0 SUMMARY

In this unit, you have learnt that correlation is the extent or degree of relationship between two variables while the index showing the degree of such relationship between the two variables is i called correlation coefficient Correlation values range from-1 to+1. Scatter-grams of different [types of relationships were shown. Pearson Product Moment Correlation Coefficient otherwise called Pearson r was also discussed with the two methods for the computation. These are the deviation and the raw score methods. The methods for calculating the Spearman rho and the Point Biserial Correlation (rpbi) were discussed in detail. The next unit will take us to the test of hypotheses to complete the module.

## 6.0 TUTOR MARKED ASSIGNMENT

i. Using any convenient correlation method, calculate the correlation coefficient of the data given below:

| S/N | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X | 31 | 24 | 50 | 12 | 45 | 47 | 09 | 51 | 25 | 60 | 15 | 10 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Y | 29 | 21 | 25 | 11 | 30 | 41 | 01 | 44 | 11 | 55 | 05 | 03 |

ii. What is the interpretation of the correlation results?

## 7.0 REFERENCES AND FURTHER READINGS

Ali, A. (1996). Fundamentals of Research in Education. Awka; Meks Publishers (Nigeria).

Anaekwe, M.C. (2002). Basic Research Methods and Statistics in Education and Social Sciences. Enugu: Podiks Printing and Publishing Company,

Ogomaka, P.M.C. (1998). Descriptive Statistics for Research Students. Owerri; Peacewise.

Olatian, S.O. and Nwoke, G.I, (19S8), Practical Research Methods in Education. Onitsha: Summer Educational Publishers.

# UNIT 5: TESTING OF HYPOTHESIS



## 1.0 INTRODUCTION

In Module one, you were introduced to the types of hypotheses. In this unit, you will learn how to test the hypotheses using some of the statistical tests.

The purpose of testing a hypothesis is to determine the probability that it is supported by facts. You may test a hypothesis by applying it to already known facts or taking it as a new appeal to experience. This same mental technique to problem-solving is also employed by science and philosophy.

Hypotheses are used as indicators of the realistic answers which researchers have to their stated problems or questions in research. So when hypotheses are tested, the results lead to establishment of new facts or confirmation of old ones. If a hypothesis is successfully verified or tested and confirmed to be true, it is then used to support a theory.

In other words, theories are developed, tested and confirmed in research through the process of hypothesis testing. This leads to the generation or advancement of knowledge. In this unit, you are going to be exposed to the rudiments of the processes involved in testing hypotheses,

## 2.0 OBJECTIVES

After going through this unit, you will be able to:

(i) explain the alpha level or level of confidence and degree of freedom;
(ii) discuss the two types of errors in hypothesis testing;
(iii) use the t-test to test a given null hypothesis;
(iv) use the relationship between correlation coefficient and t-test in hypothesis testing;
(v) use analysis of variance to test hypothesis;
(vi) use chi-square to test hypothesis;
(vii) explain the meaning of one-tailed and two-tailed tests.

## 3.0 MAIN CONTENT

### 3.1 Selection of the Level of Significance or Alpha Level

In proposing your hypothesis, you must include a confidence limit, otherwise called alpha level ($\alpha$) or significance level. Reports of inferential statistics results often include additional information such as the alpha level used and the degrees of freedom ( df ).

In most researches in education, two alpha levels are used. These are 5% (0.05) and 1% (0.01). If you choose 5% in a particular study, the implication is that if your study is replicated 100 times, the same outcome will occur 95 out of 100, and 5 out of 100 may vary due to chance. If it is 1% (0.01) level of significance, it means that if your study is replicated 100 times, you are sure 99 out of 100 will be correct while 1 out of 100 may vary due to chance factors. This is a more rigorous confidence level.

At this point, you need to note that when you test a hypothesis, you are only dealing with probability of something being true or false. Hypothesis testing helps you to make predictions ' and not to establish causation. It does not provide absolute proof. In other words, a hypothesis cannot be proved absolutely true or false.

### 3.2 Degrees of Freedom

This is the number of observations which are free to vary when certain restrictions have been placed on the data being considered. Take for instance, in your class, you ask your students to provide any four numbers which would be added to 4 to add up to 24. In this case, it is fixed, other numbers can vary. But no matter how they vary, 4 must be added to sum up to 24. Therefore, the degree of freedom here is N - 1, where N is the total number of choices while 1 is the fixed variable. As we go on, you will see the modal for getting the degrees of freedom for different tests.

### 3.3 Type I and Type II Errors

When you embark on a research study which involves the testing of hypothesis, the level of significance and the degree of freedom will enable you to take a decision about whether to accept or not to accept (reject) the hypothesis. If the null hypothesis which you have proposed is true and you accept it because your evidence supports it, then you are quite in order. It is correct. But if the null hypothesis is true based on the available evidence and you reject it, it is not correct. It is an error. Thus, the rejection of a true null hypothesis when it should have been accepted is known as Type I error.

On the other hand, if the null hypothesis is false and you accept it, instead of rejecting it, you are also not correct. In other words, the acceptance of a false null hypothesis when it should have been rejected is referred to as Type II error.

You have to note that as you try to minimize type I error by becoming too rigorous, may be you reduce the significance level from 5% to 1%, you stand the chance of making type II error by increasing the level of significance from 1% to 5%.

### 3.4 Two-tailed and One-tailed Tests

When a hypothesis is stated in such a way that it does not indicate a direction of difference, but agrees that a difference exists, we apply a two-tailed test of significance. Most of the null hypotheses are two-tailed because they do not indicate the direction of difference. They merely state that there is no significant difference between A and B. For instance, there is no significance difference in academic performance between those who went to Federal Government Colleges and those who went to State Schools. When hypothesis is stated to indicate the direction of difference, it is called a one-tailed test. For example, people who live-in high-altitude areas perform better in

long distance races. People who have stout bodies do better in short-put. Expensive cars are better in performance etc.

## 3.5 The T-test

The t-test otherwise called the student's t-test is an inferential technique. It was developed by William Gosset in 1908. There are various t-test techniques used for various tests of hypothesis concerning the following:

i. difference between population and sample means;
ii. difference between two independent samples' means;
iii. difference between matched samples' means;
iv. the significance of Pearson r;
v. difference between correlated coefficients;
vi, difference between variances that are correlated.

We are not going to treat all these in this course, hut during your master's degree programme, you will have all or most of them. For this unit, we shall take only three methods. Before we go into that, you will have to note that there are conditions for the use of t-tests. These are:

i. there must be two groups to be compared;
ii. the population from which the samples are drawn must be normally distributed;
iii. the population variances are homogenous;
iv. the samples are independent or randomly drawn from the population;
v. the variables have continuous values;
vi, suitable for both large and small samples (but not less than ten).

Note that any sample size less than 30 is regarded as small, but when the sample size is more than 30, it is regarded as large. The procedure for carrying out z-test is the same to that oft-test. While z-test is specifically used for large samples, t-test can be used for both small and large samples. When t-test is used for large samples, it approximates to z-test.

### 3.5.1 Difference between Population and Sample Means

When you want to compare a population and sample means, you will use this mode:

$$t = \frac{\overline{X} - \mu}{S\sqrt{n-1}}$$ where $\overline{X}$ = sample mean

$\mu$ = population mean
S = standard deviation
n = number.

For instance, you are given that the mean achievement score of all SS.I students in Ihitte/Uboma, in an English standardized test is 55%. A teacher conducted a study to verify this claim. He used 25 SS.I students in that locality. He drilled them on the different aspects of English syllabus for SS.I, for about eight weeks. At the end, the teacher administered the English test on the 25 students. His results are 59.85 as mean and 8.50 as standard deviation.

The first step is to propose a hypothesis (Ho). You can say the sample mean of 59.85 is not significantly grater than the population mean of 55, at an a level of 0.05 or you can say that there is no significant difference between the sample mean of 59.85 and population mean of 55.

$$t = \frac{\bar{X} - \mu}{S \left| \sqrt{n - 1}} \text{ where } \bar{X} = 59.85, \mu = 55, S = 8.50 \text{ and } n = 25$$

$$\therefore \quad t = \frac{59.85 - 55}{\frac{8.5}{\sqrt{25 - 1}}} = \frac{4.85}{8.50} \times \sqrt{24} = 0.57 \text{ x } 4.899 = 2.795$$

At this point, you have to take a decision. This will be based on the comparison of the calculated value oft-test and the value oft-test on the table or the critical region.
Now that $tc_{al} = 2.795$, df - 25 - 1 = 24, alpha level - 0.05
$\therefore$ $t_{tab}$at(25 : 0.05) = 2.060.

For decision rule, if calculated value is greater than the value in the table or critical value, Reject the null hypothesis. But, if the calculated value is less than the value on the table, Accept ho.

From this result, $t_{cal}$ is greater than the $t_{tab}$ i.e. 2.795 > 2.060. We therefore reject that there is no significant difference between the population mean and the sample, mean. It implies that there is a significant difference between the two means.

**IN-TEXT QUESTIONS**
What do you understand by the following:
(a) P < 0.05
(b) degree of freedom
(c) Type I and Type II errors
(d) Two tailed and one tailed test.

### 3.5.2 Difference between Two Independent Samples' Means

In section 3.5.1, you learnt how to find the t-test of significance when the population mean and the sample mean are given. Most of the times, you will be confronted with a situation where two samples are randomly and independently drawn from a normal population. If the variances of the samples as estimates of the population variance do not differ significantly or are homogenous, we can then say that they have a t-distribution, This is particularly when the samples' sizes are not large. Remember that a large sample size is from 30 and above. The t- statistics which you can use in this case is as follows:

$$t = \frac{\bar{X} - \bar{X}}{\sqrt{\frac{\left[(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2\right](n_1 + n)_2}{(n_1 + n_2 - 2)n_1 \times n_2}}}$$

where $S1 = \sqrt{\frac{\sum(X_1 - \bar{X}_2)}{n_1 - 1}}$

**Example 2:**

A teacher wanted to compare the academic performance of two sets of students in his school with a view to finding out whether their mean performances are significantly different. He collected samples of the two sets. His results are shown in the table below:

| Set | Mean Performance | Standard deviation | No. of Samples |
|---|---|---|---|
| 2005 | 50% | 14.50 | 80 |
| 2006 | 58% | 12.00 | 75 |

**Solution:**

i. Propose a null hypothesis $H_0$: There is no significant difference between the mean performances of the students from the two sets.

ii. $t = \dfrac{\overline{X} - \overline{X}}{\sqrt{\dfrac{\left[(n_1 - 1)S_1 + (n_2 - 1)S_2^2\right](n_1 + n_2)}{2}}}$

$$\frac{(n_1 + n_2 - 2)n_1 \times n_2}{}$$

$$t = \frac{58 - 50}{\sqrt{\dfrac{\left[(75-1)12^2 + (80-1)\left(14.5^2\right)\right](75 + 80)}{}}}$$

$$\frac{(75 + 80 - 2)75 \times 80}{}$$

$$= \frac{8}{\sqrt{\dfrac{\left[(74)144 + (79)210.25\right](155)}{153 \times 6000}}}$$

$$= \frac{(\qquad)}{\sqrt{\dfrac{10656 + 16609.75\ 155}{918000}}}$$

$$= \frac{8}{\sqrt{\dfrac{27265.75 \times 155}{918000}}}$$

$$= \frac{8}{\sqrt{\dfrac{4226191.3}{918000}}}$$

$$= \frac{8}{\sqrt{4.6036942}} = \frac{8}{2.1456221}$$

$$= 3.7285224 = 3.73$$

iii. Decision:

$t_{cal} = 3.73$, $t_{tab}$ at $(75 + 80 - 2 : 0.05/2) = t_{tab}$ at $153 : 0.05$

$t_{cal} = 3.73$. $t(153{:}0.025) = 1.96$

Since $t_{cal}$ is greater than $t_{tab}$, we reject $H_0$. It means that there is a significant difference between the mean performances of the two sets of students.

**IN-TEXT QUESTIONS**

The result of a researcher's study, to find out if there is a significant difference between the performances of males and females in his class is given below:

| Gender | Mean Performance | Standard deviation | No. of Samples |
|---|---|---|---|
| Males | 65% | 11.50 | 45 |
| Females | 58% | 14.20 | 40 |

Are the gender-wise performances significantly different?

### 3.5.3 Difference between Two-matched Sample Means

Most of the times, researchers are faced with some situations where they have to compare the performances of a set of students in two different subjects or related subjects, reaction times, tolerance levels to two types of drugs or situations etc. When this happens, the pairs of samples are not independent. The samples can be constituted through randomization. Therefore, if the samples are matched, we assume that there is no difference between the two sets of scores or variables. It implies that $\bar{X}_1 - \bar{X}_2$ . So $\bar{X}_1 - \bar{X}_2$ = d, and $\frac{\sum d}{n} = \bar{d} = 0$ .

The t-statistic is therefore given by the formula: $t = \frac{\bar{d}}{S/\sqrt{n-1}}$

where $\bar{d} = \frac{\sum d}{n}$, S = standard deviation of the ds.

**Example 3:**

A set of students took tests in both Mathematics and Statistics. Their results are as follows:

| S/N | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mathematics | 50 | 65 | 70 | 35 | 44 | 52 | 67 | 72 | 48 | 38 | 59 | 65 | 62 | 40 | 54 | 64 | 70 | 55 |
| Statistics | 48 | 60 | 74 | 30 | 40 | 50- | 69 | 70 | 50 | 42 | 60 | 70 | 60 | 29 | 52 | 61 | 70 | 53 |

Are the results significantly different?

i. Complete the table by getting d = difference (linear) between Mathematics and Statistics.

| **S/N** | **Mathematics** | **Statistics** | **D** | **$d^2$** |
|---|---|---|---|---|
| 1 | 50 | 48 | 2 | 4 |
| 2 | 65 | 60 | 5 | 25 |
| 3 | 70 | 74 | -4 | 16 |
| 4 | 35 | 30 | 5 | 25 |
| 5 | 44 | 40 | 4 | 16 |
| 6 | 52 | 50 | 2 | 4 |
| 7 | 67 | 69 | -2 | 4 |
| 8 | 72 | 70 | 2 | 4 |
| 9 | 48 | 50 | -2 | 4 |
| 10 | 38 | 42 | -4 | 16 |
| 11 | 59 | 60 | -1 | 1 |
| 12 | 65 | 70 | -5 | 25 |
| 13 | 62 | 60 | 2 | 4 |
| 14 | 40 | 29 | 11 | 121 |
| 15 | 54 | 52 | 2 | 4 |
| 16 | 64 | 61 | 3 | 9 |
| 17 | 70 | 70 | 0 | 0 |
| 18 | 55 | 53 | 2 | 4 |
| Σ | | | **22** | **286** |

ii. Find $\bar{d}$ = mean of d = $\frac{\sum d}{n} = \frac{22}{18} = 122.$

iii. Find $\Sigma d^2 = 286$.

iv. Find S - standard deviation.

$$S = \sqrt{\frac{n\sum d^2 - (\sum d)^2}{n}} = \sqrt{\frac{18 \times 286 \times 22^2}{18}}$$

$$= \sqrt{\frac{5148 - 484}{18}} = \sqrt{\frac{4664}{18}}$$

$$\sqrt{133}$$

$= \quad 16.097$

r substitute for the formula: $t = \dfrac{\bar{d}\sqrt{n-1}}{S}$

$= \dfrac{1.22\sqrt{7}}{16.097} = \dfrac{5.038}{16.097} = 0.313$

Decision: $t_{cal} = 0.313$, $t_{tab}$ at (17 : 05) - 2.131
Since $t_{cal}$ is less than $t_{tab}$ (critical value), we ACCEPT that there are no significant difference in the results, **OR** that the results are not significantly different.

### 3.5.4 Testing Hypothesis about Correlations

In the sections you have studied, you have seen how t-test can be used in different forms. You will have to note that when hypotheses testing involve the use of correlation coefficients, there are two ways to test them.

The first which you are familiar with is to use the table and find out if the correlation coefficient is significant.

The second is that, in stead of using the correlation coefficient directly from the table, you can subject it further to a t-test. In this case,

$t = \sqrt{\dfrac{1-r^2}{n-2}}$ or $t = \dfrac{\sqrt{n-2}}{1-r^2}$

**Example 4:**
A teacher wanted to find out whether students' scores in Technical Drawing have any significant relationship with their scores in Mathematics. He used the Pearson Product Moment Correlation Coefficient to do this. He came out with a correlation coefficient of r - 0.60, N = 50.
To find out if this is significant:

i. Propose a null hypothesis: The students' scores in Technical Drawing and Mathematics are not significantly related. **OR,** There is no significant relationship between the students' scores in both Mathematics and Technical Drawing.

ii. Substituting with the formula:

$$t = \frac{\sqrt{n-2}}{1-r^2} = 0.60\frac{\sqrt{50-2}}{\sqrt{1-0.60^2}} = \frac{0.60\sqrt{48}}{\sqrt{0.64}}$$

$$= \frac{4.1569219}{0.8}$$

$$= 5.196$$

iii. Find the critical value by using t(so- i: o.Q5) = 2.021.

iv. Since $t_{cal}$ is greater than $t_{tab}$ i-e. 5.196 > 2.021, we reject the null hypothesis and say the students* scores in Mathematics and Technical Drawing are significantly related.

**IN-TEXT QUESTION**
In a research study, it was found that the correlation coefficient of two variables was 0.72 and the number of the respondents was 50. Propose a null hypothesis and test it using this information at 0.05 levels.

## 3.6 Analysis of Variance (ANOVA)

In the sections earlier, you studied the t-test and its uses in verifying hypotheses. In the test for hypothesis, we can also apply the analysis of variance (ANOVA) which is referred to as Fishers Test (F-test). Analysis of variance is so called because it compares the variance (variability in scores) between the different groups (believed to be due to the independent variable) with the variability within each of the groups (believed to be due to chance). An F ratio is calculated, which represents the variance between the groups divided by the variance within the groups. A large F ratio indicates that there is more variability between the groups (caused by the independent variable) than there is within each group (referred to as the error term) (Pallant, 2011).

It is a more versatile test which can be used where two or more variables are involved for comparison. You can see that if more than two groups or variables are involved the z or t-tests cannot be used; ANOVA is used to determine the interaction effect of two or more variables, especially when the means of the sampled groups differ between and/or among the groups.

**Example 5:**
Scores of three randomly selected groups of students in an English test are given below.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| GP 1 | 15 | 20 | 12 | 10 | 9 | 7 | 6 | 11 | 18 | 14 | 5 |
| GP2 | 13 | 12 | 15 | 19 | 20 | 11 | 8 | 14 | 10 | 9 | 4 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| GP3 | 18 | 16 | 13 | 9 | 8 | 4 | 20 | 18 | 12 | 7 | 10 |

Test the hypothesis that the three groups do not belong to the same population.

| S/N | $X_1$ | $X_2$ | $X_3$ | $X_1^2$ | $X_2^2$ | $X_3^2$ |
|---|---|---|---|---|---|---|
| 1 | 15 | 13 | 18 | 225 | 169 | 324 |
| 2 | 20 | 12 | 16 | 400 | 144 | 256 |
| 3 | 12 | 15 | 13 | 144 | 225 | 169 |
| 4 | 10 | 19 | 9 | 100 | 361 | 81 |
| 5 | 9 | 20 | 8 | 81 | 400 | 64 |
| 6 | 7 | 11 | 4 | 49 | 121 | 16 |
| 7 | 6 | 8 | 20 | 36 | 64 | 400 |
| 8 | 11 | 14 | 18 | 121 | 196 | 324 |
| 9 | 18 | 10 | 12 | 324 | 100 | 144 |
| 10 | 14 | 9 | 7 | 196 | 81 | 49 |
| 11 | 5 | 4 | 10 | 25 | 16 | 100 |
| $\underline{\Sigma}$ X | **127** **11.55** | **135** **12.27** | **135** **12.27** | **1701** | **1877** | **1927** |

Find:

1. $\Sigma X_1 = \Sigma X_1 + \Sigma X_2 + \Sigma X_3$ = 127 + 135 + 135 = **397**
2. $\Sigma X^2_1 = \Sigma X^2_1 + \Sigma X^3_2 + \Sigma X^3_3$ = 1701 + 1877 + 1927= **5505**
3. $N_1$ = $N_1 + N_2 + N_3$ = 11 + 11 + 11= **33**

We shall take the correct factor to be $\frac{\left(\sum X_1\right)^2}{N_1} = \frac{(397)^2}{33}$

$\left(\sum X\right)^2$ 157609

4. Sum of squares total ($SS_t$) = $\sum X_1^2 - \frac{}{N_1}$ = 5505 − 33

= 5505 – 4776.03 = **728.97**

5. Sum of squares, between group ($SS_b$):

$$SS_b = \frac{\left(\sum X_1\right)^2}{N_1} + \frac{\left(\sum X_1\right)^2}{N_2} + \frac{\left(\sum X_1\right)^2}{N_3} - \frac{\left(\sum X_1\right)^2}{N_1}$$

$$= \frac{127}{11}^2 + \frac{135^2}{11} + \frac{135^2}{11} - \frac{397^2}{33}$$

= 1466.2727 + 1656.8182 + 1656.8182 – 4776.03

= 4779l9091 – 4776.03 = **3.891**

6. Sum of squares within groub ($SS_w$): $SS_w = SS_t - SS_b$ = 728.97 – 3.8791
= **725.09**
7. Degree of freedom, between ($df_b$) = K – 1 = 3 – 1 = **2**.
8. Degree of freedom, whiting ($df_w$) = N – K = 33 – 3 = **30**.
Where N = total number of sample
9. Variance, between groups ($V_w$) = $\frac{SS_b}{df_b}$ = $\frac{3.8791}{2}$

10. Variance, within groups ($V_b$) $= \dfrac{SS_w}{df_w} = \dfrac{725.09}{30} =$ **1.94**

= **24.17**

11. F – ratio = $\frac{V_b}{V_w}$ = $\frac{1.94}{24.17}$ = **0.08**

12. Determine the critical value of F.
From the calculation $df_b$, = 2 and $df_w$ = 30, go to the F-table and find the point of intersection of 2 and 30 at 0.05 level. This will give you the F-value i.e. 3.32.

13. Decision: F - value calculated = 0.08
F - value critical = 3.32

Since the calculated value is less than the critical value for the degrees of freedom 2 and 30, and alpha level of 0.05, we ACCEPT the null hypothesis that the scores are not significantly different.

For the purpose of presenting the result in a research report, a summary of the results is shown in a table while the computations are shown in the appendix. Thus:

| **Sources of Variation** | **Sum of squares** | **Degree of freedom** | **Variance** | **Fcal** | **F-crit** | **Decision** |
|---|---|---|---|---|---|---|
| Between groups | 3.8791 | 2 | 1.94 | 0.08 | 3.32 | Accept ho |
| Within groups | 725.0900 | 30 | 24.17 | | | |
| **Total** | **728.9691** | **32** | | | | |

**IN-TEXT QUESTION**

| S/N | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $X_1$ | 6 | 7 | 13 | 8 | 12 | 5 | 10 | 6 | 9 | 11 |
| $X_2$ | 15 | 14 | 10 | 12 | 13 | 11 | 14 | 10 | 12 | 13 |
| $X_3$ | 5 | 8 | 10 | 15 | 4 | 13 | 7 | 13 | 6 | 9 |
| $X_4$ | 10 | 7 | 5 | 8 | 9 | 8 | 6 | 4 | 7 | 3 |

Use the data above to verify a null hypothesis at 0.05.

Now that you have seen ANOVA and how to use it, we can now go to the next test. But before we do that, you have to note that ANOVA can be one-way as in the example given, two-way or multiple ANOVA. We are not going to discuss these other ones here. However, you will meet them including ANCOVA - analysis of covariance in your master's degree programme. For now, let us turn to the chi-square.

### 3.7 The Chi-Square

The word chi is pronounced kai. The chi-square is a test of independence which is used for analyzing data that are in the form of frequencies occurring in two or more mutually exclusive or discrete variables being compared. The test allows us to determine whether or not a significant difference exists between the observed frequencies of cases in each category of variables studied versus the expected frequencies or data or number of cases in each category of variables based on the null hypothesis. The observed frequency is data obtained from the actual frequency count while the expected is the data that would be obtained if equal numbers responded to the same variables equally. The larger the margin between the observed and the expected frequency counts, the higher the chi-square value. You can compare the calculated chi-square against a given critical value to determine whether it is significant. The formula for chi-square is:

$x^2 = \frac{\sum(f_0 - f_e)^2}{f_e}$ where $f_0$, is the observed frequency, and $f_e$ is the expected frequency in each cell.

**Example 6:**

A survey to determine the preference pattern of some parents on the choice of courses for their children is given in a table below. Use a null hypothesis to determine whether the indicated preference pattern is statistically significant.

| Frequency | Business | Law | Medicine | Engineering | Total |
|---|---|---|---|---|---|
| Observed | 24 | 50 | 52 | 34 | 160 |
| Expected | 40 | 40 | 40 | 40 | 160 |

Steps:

i. State the null hypothesis $H_0$: There is no significant difference between the expected and observed preference pattern of the parents at 0.05 alpha levels.

ii. Apply the chi-square formula in each cell and sum up at the end.

1. For Business = $\frac{(O-E)^2}{E} = \frac{(24-40)^2}{40} = 6.4$

2. For Law = $\frac{(O-E)^2}{E} = \frac{(50-40)^2}{40} = 2.5$

3. For Medicine = $\frac{(O-E)^2}{E} = \frac{(52-40)^2}{40} = 3.6$

4. For Engineering = $\frac{(O-E)^2}{E} = \frac{(34-40)^2}{40} = 0.9$

To take decision on the significance of the $x^2$ value, you have to find the degree of freedom df. The example discussed above is a one-variable case, so the df is given by: df = K - 1, i.e. (4 - 1) = 3. As usual, go to the chi-square table and look under df = 3, and your alpha level, which can be 0.05 or 0.01. Again, if the calculated value exceeds the value on the table, you reject the null hypothesis. In this case $x^2$ at 3 : 0.05 - 7,82. This is less than the calculated value, so we reject the null hypothesis.

Most of the times, researchers are confronted with the test for the independence of two variables. For instance, gender and opinion, or religion and choice or age and opinion. Again, each of the

variables may have two or more levels. The observed and the expected frequencies arc presented in a table called contingency table. It has a number of rows and columns.

**Example 7:**

The enrolment pattern of students to different academic programmes according to religion is given in the table below. Calculate the chi-square and test for the significance at 0.05.

| **Religion** | **Academic Programmes** | | | | |
|---|---|---|---|---|---|
| | **Business** | **Law** | **Medicine** | **Engineering** | **Totals** |
| Christianity | 50 | 35 | 48 | 45 | 178 |
| Islam | 30 | 45 | 35 | 50 | 160 |
| Traditional | 45 | 30 | 25 | 40 | 140 |
| Godian | 25 | 20 | 30 | 28 | 103 |
| **Totals** | **150** | **130** | **138** | **163** | **581** |

To solve this problem, take the table above as the table of the observed frequencies. Therefore, you will need the table for the expected frequencies. To find the expected frequency for each cell, apply the formula:

$$\frac{\text{column total} \times \text{row total}}{\text{overall total}}$$

Example, for cell 1, where the observed is 50, the expected is given by $\frac{150 \times 178}{581} = 45.96.$

For the next cell where the observed as 35, the expected is given by $\frac{130 \times 178}{581} = 39.83$ etc.

The expected frequencies are better presented in a table like the observed. See the table below

| Religion | Academic Programmes | | | | |
|---|---|---|---|---|---|
| | **Business** | **Law** | **Medicine** | **Engineering** | **Totals** |
| Christianity | 45.96 | 39.83 | 42.28 | 49.94 | 178 |
| Islam | 41.31 | 35.80 | 38.00 | 44.89 | 160 |
| Traditional | 36.14 | 31.33 | 33.25 | 39.28 | 140 |
| Godian | 26.59 | 23.05 | 24.46 | 28.90 | 103 |
| **Totals** | **150.0** | **130.01** | **138.00** | **163.01** | **581** |

To get the chi-square value, we use $\frac{\sum(O-E)^2}{E}$

Instead of taking the cells one by one, we use a table to do the same thing in a short time. Let us use a table to calculate the chi-square.

| O | E | O-E | $(O-E)^2$ | $(O - E)^2/E$ |
|---|---|---|---|---|
| 50 | 45.96 | 4.04 | 16.32 | 0.36 |
| 30 | 41.31 | -11.31 | 127.92 | 3.10 |
| 45 | 36.14 | 8.86 | 78.50 | 2.17 |
| 25 | 26.59 | -1.59 | 2.53 | 0.10 |
| 35 | 39.83 | -4.83 | 23.33 | 0.59 |
| 45 | 35,80 | 9.20 | 84.64 | 2.36 |
| 30 | 31.33 | -1.33 | 1.77 | 0.06 |
| 20 | 23.05 | -3.05 | 9.30 | 0,40 |
| 48 | 42.28 | 5.72 | 32.72 | 0.77 |
| 35 | 38.00 | -3.00 | 9.00 | 0.24 |
| 25 | 33.25 | -8.25 | 68.06 | 2.05 |
| 30 | 24.46 | 5.54 | 30.69 | 1.25 |
| 45 | 49.94 | -4.94 | 24.40 | 0.49 |
| 50 | 44.89 | 5.11 | 26.11 | 0.58 |
| 40 | 39.28 | 0.72 | 0.52 | 0.01 |
| 28 | 28.90 | -0.90 | 0.81 | 0.03 |
| | | | | **14.56** |

From the calculation shown above, the calculated value is:

$x$ = 14.56, df = (c - 1) (r - 1) - (4 - 1) (4 - 1) - 9.

For decision, go to the table to look for the critical value at df = 9, alpha = p = 0.05. X 2tab - 16.92.

Since the calculated value of 14,56 is less than the critical value of 16.92, we Accept the null hypothesis that there is no significant difference between the observed values and the expected values.

**IN-TEXT QUESTION**

Use the data below to verify your proposed null hypothesis:

| **Gender** | **VX** | **VY** | **VZ** | **Total** |
|---|---|---|---|---|
| Male | 55 | 40 | 50 | 145 |
| Female | 35 | 25 | 40 | 100 |
| Total | 90 | 65 | 90 | 245 |

## 4.0 CONCLUSION

Now that you have successfully worked through this unit on how to test hypotheses, you are now prepared to carryout your research project work. But before you go properly into that, we shall introduce you to how to write research reports in the next unit.

## 5.0 SUMMARY

In this unit, we have discussed the selection of the alpha level or significance level and we said the two most common alpha levels used in research are 0.05 and 0.01. We touched upon the degrees of freedom, Type J error and Type II error as the likely errors that can be made in decision making in the test of hypothesis.

Hypotheses can be frame in two formats, which are directional and non-directional. This implies that we have two types, vis-a-vis one tailed test and two tailed test. You also studied the

different types of tests used in testing hypotheses. The t-test, the F-test and the chi-square are the prominent.

In the next unit, you will be introduced to how to write your research reports.

## 6.0 TUTOR MARKED ASSIGNMENT

A class of students did a test in Introduction Technology when they were in JS.2. The same class of students studied Technical Drawing in their SS.2. The results are given in the table below. What is the correlation coefficient of these sets of scores? Propose a null hypothesis and verify it using t-test on the result of the correlation coefficient.

| S/N | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Introduction Technology | 20 | 18 | 17 | 25 | 22 | 15 | 13 | 10 | 19 | 24 | 16 | 8 | 5 | 14 | 12 |
| Technical Drawing | 25 | 20 | 18 | 24 | 20 | 17 | 18 | 15 | 19 | 20 | 20 | 12 | 10 | 22 | 14 |

## 7.0 REFERENCES AND FURTHER READINGS


Ali, A. (2015). Fundamentals of Research in Education. Awka: Meks Publishers (Nigeria),

Anaekwe, MC. (2015). Basic Research Methods and Statistics in Education and Social Sciences. Enugu: Podiks Printing and Publishing Company.

Denga, D. and Ali, A. (2014). An Introduction to Research Methods and Statistics in Education and Social Sciences. Jos: Savannah Publishers Limited.

Julie Pallant. (2016). *SPSS Survival Manual: A step by step guide to data analysis using SPSS.* (4th edition). China: Everbest Printing Co

.

# UNIT 6: RESEARCH REPORTS WRITING



## 1.0 INTRODUCTION

The final stage of any research process is the writing of the research report. Research is very important, because the findings generated can be used for rational decision-making and, according to Nkpa (1979), as a springboard for further research. The main aim of writing research report is to communicate or disseminate the research findings to the literate audience. In writing the research report, the impersonal mode is preferred. That is to say, instead of say "I did this", you should say "the study was carried out to do this".

You will have to note that in presenting a research report, you have to use the required format. Most institutions have their own format. These formats or house-styles do not vary significantly from the general format. National Open University of Nigeria (NOUN), School of Education, has its own house-style. You supervisor will let you have it. For the purpose of this unit, we shall discuss the general format. According Clark and Creswell (2015) a good research report, the researchers tend to address a common set of topics when interpreting their research results. These interpretations include the following elements:

- a summary of the major results,
- a discussion relating the results to the literature,
- personal reflections of the researcher about the meaning of the research,
- implications for practice,
- limitations of the study,
- future research needs, and
- the overall significance of the study.

## 2.0 OBJECTIVES

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

(i) Itemise the chapter titles and sub-titles in a research project report;
(ii) Prepare a research report based on the given format.

## 3.0 MAIN CONTENT

### 3.1 Sample Format of a Research Report

As you have seen in the introduction, a research project report is a detailed account of what the researcher has done in the process of carrying out the research the findings of this study. The report is not presented in any form. It follows an agreed format as summarized below. This format is only a guideline. Though this is the conventional format, only relevant section should be used in line with your house-style.

**1. Preliminary pages:**
- i. Title page
- ii. Approval / Acceptance page

**2. Certification page**
- i. Dedication
- ii. Acknowledgement page
- iii. Abstract
- iv. Table of Contents
- v. List of tables
- vi. List of figures
- vii. List of appendices

**3. Chapter 1: Introduction**
- i. Background to the Problem
- ii. Statement of the Problem
- iii. Purpose / Objectives of the Study
- iv. Significance of the Problem
- v. Scope of the Study
- vi. Research Questions and/or Hypotheses
- vii. Definitions of Terms

**4. Chapter 2: Literature Review**
- i. Review of Related Literature
- ii. Conceptual Framework

**5. Chapter 3: Research Methodology**
- i. Research design
- ii. Population
- iii. Samples and Sampling techniques
- iv. Instrumentation - construction of instruments, validation, reliability of instruments, administration and scoring
- v. Methods of data analysis

**6. Chapter 4: Presentation of Results**
- i. Data analysis and findings
- ii. Summary of major findings

**7. Chapter 5: Discussion**
- i. Interpretation of findings
- ii. Discussion of findings
- iii. Implication of the study
- iv. Recommendations
- v. Limitations
- vi. Suggestions for further study

**8. Supplementary page:**
- i. Bibliography
- ii. Appendices

iii. Index

## 3.2 Steps in Research Report Format

You have already noted that a research report is a straight forward, clearly and precisely written document in which you attempt to explain how you have resolved the problem before you. The presentation, in this unit, is consistent with the most acceptable formats. So let us explain them.

### 3.2.1 Preliminary Pages

***i.*** ***The title page:*** This is the first page of this section. It contains the title of the study, the name of the author, the relationship of the research to a course or degree requirement, name of the institution where the report is to be submitted, and the date of presentation.

The title should be concise and state clearly the purpose of the study. The essential elements to be included in the title are the major variables and the target population. These should be phrased in such a way as to describe what the study is all about. You should not state your title so broadly that it may claim more than it can actually deliver. For instance, sex differences in the enrolment of SSCE candidates in Technical Drawing from 2004 to 2007, or the effect of group discussions on learning outcomes in the Open and Distance Education system. You can note the variables here. The title should be typed in capital letters, single-spaced, and centered between the right and left margins of the page.

***ii.*** ***Approval/Acceptance page:*** The specifications vary from institution to institution. It contains some of the following information; the names, signatures of the head of department, the dean, the supervisors) and dates, the names(s) of the student(s).

***iii.*** ***Certification page:*** This contains the attestation of originality of the research project. It may also include the name and signature of the external examiner.

***iv.*** ***Dedication:*** Here, emotionally-lad en words may be permitted in order to pay tribute to persons who are dear to the author or those who contributed in one way or the other to the success of the project and those who would particularly be interested m the research findings.

***v.*** ***Acknowledgement page:*** This is used to express gratitude to those who helped in the process of conducting the research and preparing the report. It should be simple and restraining.

***vi.*** ***Abstract:*** This is a succinctly summarised form of the report containing the aim of the investigation, the sample, methods of investigation, the instruments used for data collection, the analysis and findings.

***vii.*** ***Table of Contents:*** This serves an important purpose of providing the outline of the contents of the report. It lays out in a tabular form, the chapters, headings and sub- headings of the report. It is sequentially arranged and numbered from the preliminary to the supplementary pages. Page references for each topic are so indicated.

***viii.*** ***List of tables and figure and appendices:*** If tables and/or figures are used in the report, a separate page is included for each list. It should indicate the page numbers in which the tables or figures presented in the report are located. The numbers and titles are serially listed. Also contained is the list of appendices that are embodied in or annexed to the report. The pages of the preliminary section are numbered with lower-case Roman numerals (i, ii, iii, iv, v, etc).

### 3.2.2 Introduction

i. ***Background to the Problem:*** Here, such factors or circumstances that informed the investigation are traced. It is presented using reasoned statements to show that it is worthwhile to dissipate resources to carryout the study. It shows the nature, scope, and current status of the factors affecting the problem. It has to be presented in a way as to be clear and convincing to the reader.

ii. ***Statement of the Problem:*** The problem as already highlighted is stated in simple, clear and unambiguous terms. This is not required to be too long.

iii. ***Purpose of the Study/Objectives of the Study:*** These go interchangeably, but it states the specific aspects of the study and the reasons for focusing on them. It includes statements of what should be accomplished and all that would be investigated.

iv. ***Significance of the Problem:*** The usefulness, the utility value of the research or findings of the research should be articulated. The institutions, groups or individuals who are expected to profit or benefit and the benefits expected to accrue to them are to be stated in this section.

v. ***Scope of the Study:*** This is interchanged with the delimitation of the study. Here, you will have to indicate the extent to which the study will be covered. It involves the geographical area, time period, and variables to be covered.

vi. ***Research Questions and/or Hypotheses:*** These are formulated based on the type of research and the variables under investigation. They should be formulated to provide answers to the problems under study.

vii. ***Definitions of Terms:*** The essence of definition is to ensure that the reader understands the specific meanings ascribed to the terms by the author So you have to use this to educate the readers on the operational meaning of any coined, technical words, phrases or expressions which cannot otherwise be understood because of their unconventional usage.

### 3.2.3 Literature Review

i. ***Review of Related Literature:*** This is the second chapter of your project report. It is meant to give the reader an understanding of some of the works or study already carried out in the area of the project. It will also give the reader an overall picture of the problem you are solving. You are therefore required to review only the important literature related to your study, abstract previous research studies and review significant writings of authorities in the area under study.

By so doing, a background for the development of your study will be provided. It will also bring the reader up-to-date. Apart from providing evidence of the investigator's knowledge of the field of study, it highlights the areas of agreement or disagreement in findings or gaps in existing knowledge.

Do not use the article-by-article presentation in your literature review. In other words, do not collect abstracts of previous researches and string them together without any attempt at continuity or logical organization. Again do not make excessive use of quotations. Quotations are used only when the material quoted is especially well written and can be inserted without spoiling the continuity of the presentation (Olaitan and Nwoke, 1988).

ii. ***Conceptual Framework:*** This states the concept that informed the study. These concepts such as system concept, management by objectives concept, etc. will assist you to bring out salient points that would assist to important literature related to your study, abstract of previous research studies and review significant writings of authorities in the area under study.

### 3.2.4 Research Methodology

i. ***Research design:*** This lays out the master-plan for the research project. It shows the extent to which extraneous variables were controlled or eliminated. You should therefore describe any plan used clearly, even if it cannot be classified under a conventional label. All lapses should be reported as a limitation.

ii. ***The Population:*** You should specify all the necessary parameters to ensure that the constituents and characteristics of the target population are unambiguous. The target population may be people, animals, objects or events.

iii. ***Samples and Sampling techniques:*** The size of the sample and how the sample was selected should be so described in such a way as not to leave the reader in doubt about what you have done. Do not just say 100 respondents were randomly selected from the population. Specify the method in which the simple random sampling was used. Is it by the use of table of random numbers, describe whether pieces of numbered papers were jumbled in a box and picked up at random, etc.

iv. ***Instrumentation:*** In this section, you have to describe in full details the tools for data collection. Such tools like questionnaire, attitude scales, tests, etc. should be fully described to show their characteristics. You will have to report the reliability indices and validation procedures. Where you used a standard instrument, in your report, you have to give the rationale for the appropriateness. Where a new instrument is developed, you have to outline the necessary procedures followed in both the construction and validation,

v. ***Data Collection:*** What methods did you use in your data collection? Did you use research assistants? If yes, did they undergo training? Did you collect the data personally, or by post? What problems did you encounter in the process of data collection? All the steps which you have taken to ensure the collection of valid that should be reported.

vi. ***Methods of data analysis:*** In this section, you will describe the techniques which you applied in the data analysis and the reasons for the choice. The reasons may be in relation to the type of design, nature of the samples on the type of data. Try to use the simplest, well known method of data analysis. But where you use a mode of analysis not widely known details of such method should be reported.

### 3.2.5 Results and Discussion

i. ***Presentation and Analysis of data:*** This is the heart of the research report. The results are clearly and concisely set out using the most illuminative modes of presentation. Tables, figures, graphs and textual descriptions are used to clarify significant relationships. They should be serially numbered and titled so as to be self explanatory. They should be simple and should be directly related to the hypotheses and/or the research questions.

ii. ***Interpretation of the finding:*** The most important task which you have to undertake in writing the results of your study is to identify and interpret the major findings. You should be able to discuss possible reasons why the results occurred the way they did. You should try to fit them into the findings of previous research, suggest the applications to the field and make theoretical interpretations.

### 3.2.6 Summary and Conclusions

i. ***The Summary:*** In this section, you should clearly and concisely restate the problem, the hypotheses and/or research questions, the main features of the method omitting most of the details concerning the subjects and measures and list the main findings.

The summary must be very brief, but consistent with a clear presentation of all important information about the problem, method and findings. The findings should be listed by number. You should summarize each major finding in one or two statements.

ii. ***The Conclusion:*** This gives answers to the questions raised or the statements of acceptance or rejection of the hypotheses. It should be based solely on the findings generated by the research.

iii. ***Implication of the study:*** In this section, you may include ideas on the relevance of the findings to educational theory and practice. But these ideas should be directly be derived from the study.

iv. ***Suggestions for further study:*** It may be appropriate here to suggest areas of problems for further investigation. This is made as a result of matters arising from the research.

**3.2.7 Supplementary Pages**

i ***Bibliography:*** In this section, you should include all references cited in the report and those not cited, but consulted to shed light on the problem, References are cited uniformly and according to a given style.

Most universities adopt the APA format. References are done serially and alphabetically. You can look for the APA format and go through it.

ii ***The Appendices:*** This contains extra information which is part of the report the reader should know about, but not necessarily for inclusion in the main report. They include long tables, forms, instruction aids, data collecting instruments, items analysis data, scoring protocols and procedures, lengthy quotations etc.

Each separate entry heading is listed as APPENDIX A, APPENDIX B, etc.

**Self Assessment Exercise**

Go to any university library and select three different research projects. List the items on the table of content and compare them.

## 4.0 CONCLUSION

At the end of your programme, you are expected to carryout a research. At the end of the research, you are also expected to submit a written report of the investigation. In this unit, you have gone through the involvement in the writing of the report. A very important demand here is that you must be as objective as possible in your report. At the initial stage, you cannot make any statement that would show you are in favour or against an idea. Your report should be devoid of emotional or subjective statements. You should arrange the different parts of the report so as to make it possible for a reader to easily locate any section of particular interest to him,

## 5.0 SUMMARY

In this unit, we have discussed and presented a sample format of a research report. We have also discussed these steps in details stating from the preliminary stages to the supplementary stages. We have emphasized that your reports should not be presented with personal pronouns like I, my, we etc. Instead use impersonal pronouns and passive voice. You should make sure that the report is written in a clear, simple and straightforward style. Your motive should be effective communication. Therefore, use very simple language. You should always be brief so as not to bore your reader. Abbreviations should only be used after they have been written in full earlier. Avoid the use of generalizations or conclusions, which are not supported by the findings, We also said that every source cited in the work or used but noted

cited in the work should be documented in the reference page. Improper citation or inability to give details of a source cited in the body of the work should be documented in the reference page. Improper citation or inability to give details of a source cited in the body of the work should be avoided. Remember that proofread the report thoroughly after typesetting. This will help you not submit avoidable errors.

Congratulations for being part of the success story of NOUN, and for graduating in this programme.

## 6.0 TUTOR MARKED ASSIGNMENT

Pick up any four (4) research projects. Study the abstracts. What are the things that are common to all of them?

## 7.0 REFERENCES / FURTHER READINGS


Ali, A. (2015). Fundamentals of Research in Education. Awka: Meks Publishers (Nigeria),

Anaekwe, MC. (2015). Basic Research Methods and Statistics in Education and Social Sciences. Enugu: Podiks Printing and Publishing Company.

Denga, D. and Ali, A. (2014). An Introduction to Research Methods and Statistics in Education and Social Sciences. Jos: Savannah Publishers Limited.

Julie Pallant. (2016). *SPSS Survival Manual: A step by step guide to data analysis using SPSS.* (4th edition). China: Everbest Printing Co