NATIONAL OPEN UNIVERSITY OF NIGERIA

FACULTY OF AGRICULTURAL SCIENCES

Department of Agricultural Economics and Extension

Programme: Hospitality and Tourism Management

Course Code: HTM403

Course Title: RESEARCH METHODS IN HOSPITALITY AND TOURISM

Credit Unit: 2 Units

ISBN: 978-978-058-181-7

HTM403

RESEARCH METHODS IN HOSPITALITY AND TOURISM

Course Developer/Writer: Dr. Esheya Samuel Esheya

Department of Agricultural Economics and Extension,
Faculty of Agricultural Sciences,
National Open University of Nigeria.

Course Content Editor: Prof. Ezike, K.N.N.

Department of Agricultural Economics, Management and Extension,
Ebonyi State University, Abakaliki.

Year Developed/Written: June 2022.

National Open University of Nigeria

National Headquarters

91, Cadastral Zone, Nnamdi Azikiwe Express Way, Jabi, Abuja
Nigeria.

E-mail: centralinfo@nou.edu.ng

URL: www.nou.edu.ng

Published by

National Open University of Nigeria

Printed ……..

ISBN: …….

CONTENTS PAGE

Introduction

The course, Research Methods is a core course, which carries two (2) credit units. It is prepared and made available to all degree course students offering Hospitality and Tourism related Programme in the Faculty of Agricultural Sciences**,** Department Economics and Extension at the Nation Open University of Nigeria. Research Methods is the application of research principles to the operations of tourism and hospitality industry. This course material is useful in your academic pursuit as well as in your workplace as managers and administrators.

What You will Learn in this Course

This course consists of five modules which are sub-divided into 14 units. This course guide tells you what the course is all about. What course materials you will be using and also suggests some general guidelines for the amount of time you are likely to spend on each unit of the course in order to complete it on schedule. It also gives you guidance in respect of your Self- Assessment Exercises (SAEs) which will be made available in the assignment file. Please attend those tutorial sessions. The course will introduce you to the rudiments of research methods.

Course Aim

The main aim of this course is to arm you with adequate information on the concept of research and methodology in hospitality and tourism management. The course also aims at making you have a greater understanding of the fundamentals of research and statistics as applicable to hospitality and tourism management. This will prepare the student for a future career in hospitality and related disciplines.

Course Objectives

To achieve the aim set out, the course has a set of objectives which are set out as intended learners' outcome under each unit. You should read these objectives before you study the unit. After going through this course, you should be able to:

- Discuss the meaning of research;
- Evaluate the research methods;
- Describe the importance, need and significance of research.
- Discuss the major concepts used in research;

- Write short notes on other terms used in research methodology.
    - Discuss the classification of research based on purpose;
    - Analyse research classification based on method and evidence required;
    - Describe the classification of research on the basis of method of data analysis.
    - Discuss research design in various ways;
    - Analyse the properties of a good research design;
    - State the relevance of research design.
    - Distinguish between population and sample
    - Analyse the concept of sampling.
    - Define sampling method or technique;
    - State the types of sampling methods with relevant examples;
    - Analyse the different categories of sampling methods/techniques.
    - Discuss the research process.
    - Analyse the sequential steps or stages of the research process;
    - State the roles of research in hospitality and tourism management.
    - Explain the meaning of research planning;
    - Write short notes on research execution,
    - Analyse the steps research planning and execution.
    - Discuss the meaning of results and data collection;
    - Analyse the steps in data reduction and organization;
    - Evaluate the statistical tools for data analysis.
    - Discuss the descriptive statistics used for data analysis;
    - Analyse statistical information on graphs;
    - Interpret research results based on analysed data.
    - Discuss the meaning of inferential statistics;
    - Demonstrate examples of inferential statistical tools;
    - Evaluate the application of these tools to data analysis and interpretation.
    - Discuss research proposal;
    - Analyse the steps in preparing a research proposal;
    - Describe the components of a research proposal.
    - Discuss research report;

- Write an essay on how to prepare a research report;
- Identify the target audience of a research report;
- Evaluate the functions of a good research report.
- Discuss the roles of statistics in research; and
- Evaluate the computer software applied to research.

Working through the Course

This course involves that you devote a lot of time to read and study the contents. Each unit contains self-assessment exercises for this course and at certain points in the course you would be required to submit assignments for assessment purposes. At the end of this course, there is a final examination. I would therefore advice that you attend the tutorial sessions where you would have the opportunity of comparing knowledge with your colleagues.

Course Materials

You will be provided with the following materials

· Course guide

· Study units

· References

· Assignments

· Presentation schedule

STUDY UNITS

There are five modules of 14 units in this course, which should be studied carefully.

Assessment

There are two components of assessment for this course:

· The Tutor Marked Assignment (TMA)

· The end of course examination.

Tutor-Marked Assignment

The TMA is the continuous assessment component of your course. It accounts for 30% of the total score. You will be given four TMA's by your facilitator to answer before you can sit for the final examination.

Final Examination and Grading

This examination concludes the assessment for the course. The examination will account for 70% of total score. You will be informed of the time for the examination.

Summary

This course intends to provide you with underlying knowledge of research methodology principles for the study of Hospitality Management and Tourism.

TABLE OF CONTENTS **PAGE**

HTM403

RESEARCH METHODS IN HOSPITALITY AND TOURISM

MAIN CONTENT

MODULE 1 CONCEPT AND GOALS OF RESEARCH

Unit 1: Meaning, Characteristics and Significance of Research

UNIT 2 Definition of Concepts Used in Research

UNIT 3 Types and Classification of Research

Unit Structure

1.1 Introduction

Our environment is dynamic and its problems are enormous. This requires that complex decisions must be made in systematic procedures as much as possible with the aid of research. Research is thus an activity that leads us to finding new facts, information, assisting us in verifying the available knowledge and in making us question things that are difficult to understand as per existing data. *To be successful manager in hospitality and tourism management and related disciplines, it is important for you to know how to go about making the right decisions by being knowledgeable about the various steps involved in finding solutions to problematic issues through empirical research*. Many writers and scholars view the topic 'research methods' from different angles and perspectives. While some perceive it as methods of inquiry, others view it as a process of gathering information and making deductions from it. Research methods seek through methodological process to add to one's knowledge and hopefully to others, by the discovery of non-trivial facts and insights. Thus, research does not imply the routine application of what is already known. Rather, research is oriented towards the discovery of the relationships existing among the phenomena of the world. It is reserved for procedure designed to discover facts and relationships that will bring about more effective knowledge. The aim of science is to produce knowledge and to explain some aspects of the world (real world phenomena) around mankind. Despite the disciplinary differences between the different fields of scientific study, scientific knowledge regardless of the discipline, shares unique defining attributes which hospitality and tourism management cannot be an exception.

1.2 Learning Outcomes

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

- Discuss the meaning of research;
- Evaluate the research methods; and
- Describe the importance, need and significance of research.

1.3 Meaning, Characteristics and Significance of Research

Research compliments exiting knowledge rather than simply leading to an increase in one's understanding of previous work. According to Chukwuemeka (2002), research is simply the process of arriving at dependable solution to problem, through the planned and systematic collection, analysis and interpretation of data. Research is reserved for activities involved in the discovery of facts and relationships that make knowledge more effective. The challenges of economic development in Nigeria and Africa in general calls for intensified research to foster ideas and sound reasoning for growth and prosperity. Research is founded on the scientific method and the purpose scientific method is to obtain reliable knowledge about nature or reality by verifiable

means. This implies that the two pillars of science are logicality and observation. Scientific information must show sound logic and must correspond with what is observed (Eboh, 2009).

1.3.1 Meaning of research

Research can be described in several ways but it is a simple concept within the reach of every scholar. Research is defined as the process of arriving at dependable solutions to problems through the planned and systematic collection, analysis and interpretation of data. It is an objective, systematic, controlled and critical activity, planned and directed towards discovery and development of dependable knowledge. Thus, research must be objective to arrive at the truth. This is because the existence of truth is independent of man's knowledge and opinion which can only be discovered through empirical and objective process. As a systematic process, research follows a replicable orderly set of methodological procedures with built-in checks at each step to ensure objectivity. While other methods of searching for the truth like experiencing, reasoning and speculating are subjective and have no obvious and replicable order, research uses carefully planned reproducible observation and investigation to enable a deep and full understanding of the true nature of phenomena. This is painstaking, logical and systematic process.

Research as a critical study is closely tied to its objective nature. A research is a process of critical thinking in that judgment is reserved until empirical evidence proves our speculations or hypothesis right. Scientific knowledge evolves critical solution to problem during which man only proposes solutions, hypothesis, but nature disposes of their truth of falsity. The method used must be such that can stand both internal and external criticism. Research is a rational process, which tries hard to overthrow its speculations than to defend them (Iyiogwe, 2002). The aforementioned

qualities delimit the kinds of activities and processes which can be considered as research regardless of where it is carried out, be it in the physical, social, and behavioural sciences. Research involves the same processes and its results are subject to the same standard. Research could be regarded as the process of arriving at dependable solutions to problems through the planned and systematic collection, analysis, and interpretation of data. It is the most important tool for advancing knowledge, for promotion of progress and for enabling man to relate more effectively to his environment, to accomplish his purpose and to solve his conflict.

The motivating factor behind every research is the existence of a problem and the urge to get it resolved. Without a problem-situation, there will be no research. Thus, it is the existence of the problem that necessitates the search for logical solution and hence, research. Research is also defined as the creation of new knowledge and the use of existing one in a new and creative way so as to generate new concepts and methodologies which may include synthesis and analysis of previous studies to create new outcomes. Research is a human activity based on intellectual application in the investigation of matter. The primary objective of research is discovering, interpretting and the development of methods and systems for the advancement of human knowledge on a vriety of scientific of our world and the universe. It is a systematic investigation designed to develop and contibute to generalize knowledge about the variable one is interested. As an art of scientific investigation, research means a scientific and systematic search for knowledge and information on a specific topic.Research comprises of defining and redefining problem, formulating hypothesis, collecting, organising, and evaluating data, making deduction and reaching conclusion and at least carefully testing the conclusion to determine whether they fit the formulating hypothesis or not (Cifford Wooay, 2020).

Thus, research is a systematic, logical and structured, with specified steps in specified sequence in accordance with a well-defined set of rules." Research is an investigative process of finding reliable solution to a problem through a systematic selection, collection, analysis and interpretation of data relating to the problem. In other words, research is all activities that makes us discover new knowledge about things around us in the universe. Research is a creative and systematic work undertaken to increase the stock of knowledge. It involves the collection, organization and analysis of information to increase understanding of a topic or issue. A research project may be an expansion on past work in the field. To test the validity of instruments, procedures, or experiments, research may replicate elements of prior projects or the project as a whole.

Research in simplest terms is searching for knowledge and searching for truth. In formal sense it is a systematic study of a problem attacked by a deliberately chosen strategy which starts with choosing an approach to preparing blue print (design) acting upon it in terms of designing research hypotheses, choosing methods and techniques, selecting or developing data collection tools, processing the data, interpretation and ending with presenting solution/s to the problem.

Another definition of research is given by John W. Creswell (2014), who states that "research is a process of steps used to collect and analyze information to increase our understanding of a topic or issue". It consists of three steps: pose a question, collect data to answer the question, and present an answer to the question.

The Merriam-Webster Online Dictionary defines research in more detail as "studious inquiry or examination; *especially* : investigation or experimentation aimed at the discovery and interpretation of facts, revision of accepted theories or laws in the light of new facts, or practical application of such new or revised theories or laws". Above all, the basic goal of the research process is to produce new knowledge or deepen understanding of an existing one.

The major steps in conducting research are:

- Identification of research problem
- Literature review
- Specifying the purpose of research
- Determining specific research questions
- Specification of a conceptual framework, sometimes including a set of hypotheses[32]
- Choice of a methodology (for data collection)
- Data collection
- Verifying data
- Analyzing and interpreting the data
- Reporting and evaluating research
- Communicating the research findings and, possibly, recommendations.

1.3.2 Definition of Research Methods

Method simply means a particular procedure for accomplishing or approaching an issue, especially an existing, systematic or established one. The steps involved in finding responses to the research questions comprise research methodology. Therefore, methodology can be defined as a set of specific procedures or techniques used to identify, select, process, and analyze information about a topic. It is the systematic, theoretical analysis of the methods applied to a field of study. It comprises the theoretical analysis of the body of methods and principles associated with a branch of knowledge. Before the research begins it is important to decide the methods of data collection, the data could be qualitative or quantitative or both. The process of research addresses two major questions i.e. what is to be found and how it is be found (Ranjit, 2014). Research methods or methodology is the systematic process or procedure designed for generating, collecting collating

and analysis of data required for solving a specified research problem. Research methodology consists of the following essential parts:

a. Scope of study and area of coverage
b. Sources of data
c. Population of the study
d. Determination of the sample size
e. Sampling procedure/sampling design/sampling plan
f. Methods of data analysis
g. Administration of research instrument

a. Scope of study and area of coverage

This is used to define the jurisdiction of the study in terms of time, demographical and geographical elements, with respect to the research problem and the associated variables stated earlier. It also decides how (interview, observation, mailed questions, etc.) the data is collected.

b. Sources of data

There are basically two sources of data used for research. These are primary (original) data and secondary (documented) data. This section describes the place, people or objects from which each data sources are obtained.

c. Population of the study

Population is defined as the total number of people, objects, events and things that possess the required characteristics of interest for a specified study. Population represents the target of the study for data collection. This can be large or small depending on the nature of the study but it is

necessary to define a study's population and specify its elements in order to make it complete and useful.

d. Determination of the sample size

Sampling enables the researcher to use the information obtained from part of the population to take decision on the whole census. However, obtaining information from all members/elements of a population is generally more accurate than sampling. But without sampling, research involving very large or infinite population would be expensive and time consuming if not impossible. However, deciding the appropriate size of sample for study is a major problem for some researchers. The larger the sample size, the more accurate the information derivable from it, and the higher the costs of obtaining the required information. The size of the sample must consider the desired accuracy of the sought information and available resources in terms of time and research funds. Many methods are available for determining sample size which are acceptable to all researchers but Taro Yamani (1964) formular is the most commonly applied.

Example: If we assume that the National Open University of Nigeria has the following staff composition: senior staff = 211, intermediate staff = 428, and junior staff = 622. You are expected to determine the sample size using Taro Yamani formular.

Solution: According to Taro Yamani (1964), the formular for determining a sample size from a population of study is given by:

$n = N/1+(Ne^2)$

where:

n = desired sample size

N = total population

e = acceptable error limits (± 0.05).

Here N = 211 + 428 + 622 = 1261

$e = 0.05$

$e^2 = 0.0025$

$n = 1261/1 + 1261(0.0025)$

$= 1261/1 + 3.15$

$= 1261/4.15$

$n = 304.$

So, 304 should be the appropriate sample size for an individual researching on a population of 1261 people.

e. Sampling Procedure/ Sampling techniques

This specifies the methods by which each of the units will be selected for measurement or collection of information. The basic methods of sampling are probability and non-probability. Here statement is to be made on how the chosen method was employed in the study for the benefit of researchers that may be interested in related areas. The two basic alternatives of sampling techniques are:

i. probability sampling

ii. non-probability sampling

i. Probability Sampling

The variants of probability samples include: simple random sampling, stratified random sampling, systematic sampling and cluster sampling. A probability sample is one in which the selection of

the sampling units is based on chance, and for which the chance of each unit being selected is known and this makes it possible for a sampling error to be calculated. **Simple random sampling** is the type of sampling method in which every member of the population has an equal and independent chance of being selected in the sample to be drawn for the study. The selection of one element does not affect the chances of another element being included in the sample. **Stratified random sampling** is a method of sampling where the population from which samples are to be drawn is divided into groups called strata on the basis of certain characteristics common among each group. A stratified sampling technique is one that recognizes different groups or the composite measures to include them in the sample. **Systematic sampling technique** involves taking every unit or element in the frame, after picking a random starting point. It involves selecting every nth element from the target population. The nth element represents the desired respondents that should be included in the sample chosen in a sequential order hence the term systematic. The nth elements are chosen from a list containing the members of the population.

Cluster sampling is a process in which the population is divided into subsets called clusters which contain two or more of the population elements. A sample is randomly selected from among the clusters and the members or elements therein are studied. Cluster or area sampling involves selecting members of a sample in groups rather than individuals.

ii. Non- Probability Sampling

Non- probability samples are drawn without giving all the elements equal and fair chances of being selected. The criteria for being included in the sample depend mainly on the judgement of the researcher. Again, non-probability sampling is one in which the likelihood of every member of a

target population being included in the sample is not known. Non-probability sampling includes quota sampling, judgement or purposive, volunteer and convenience sampling.

The term quota sampling implies a predetermined number of what is desired. For instance, a study that has gender bias might insist on including a certain number of women and men. Besides, quota sampling is one in which the elements are arbitrarily picked in correspondence to the known proportions of the different groups into which the population has been subdivided. **In judgment or purposive sampling** the researcher selects elements from the population on the basis of his presumption that they are suitable for the study or representative of the population. Purposive sampling is very useful in certain research situations when the goal of generalizing on the population is not needed. **Volunteer sampling** is a method in which the researcher may request people to volunteer and can only base the study on the volunteers. The researcher takes his sample from among those elements in the population that willingly offer themselves for the application of his research instrument. This makes the results of studies using information obtained from volunteers biased and unacceptable to some critics because the attitude of such volunteers may be at variant from the general behavior pattern of the entire population. **Convenience sampling** entails selecting elements from a population to constitute a sampling, based on the convenience of the researcher, without any serious regard to their representatives.

f. Methods of data analysis

Data analysis is done descriptive or inferential statistics or both. Descriptive statistics which include tables, percentage, bar charts, pie charts, histograms, frequency polygons, means or averages, modes, medians and variances are used for summarizing the data presented while inferential statistics such as chi-square, t-test, regression analysis, correlation analysis, analysis of

variance etc. are used for the estimation of population values, evaluation of the relationship between two or more variables and testing of hypotheses.

g. Administration of research instrument

After a researcher has done his proposal and selected his sample, the next thing is for him to collect data required for his topic of study. Data collection is the greatest effort of the researcher which could make or mar the research. The researcher is interested in primary and secondary data. Primary data are called field data and the important methods of collecting primary data are observation, interview, telephone and panel surveys. A questionnaire is a list of questions or statements which require respondents to make responses. It is a data gathering instrument in which respondents are given standard or uniformed questions. Questionnaires could be administered directly or indirectly to respondents. Steps in questionnaire preparation include:

i. The address of the researchers should be written on the top-right side of the questionnaire;
ii. The purpose for which the research is being conducted should be stated;
iii. A guaranty of security of the respondents should be included in the introductory part of the questionnaire;
iv. Economics of words and space when structuring questionnaire should be adopted; and
v. The questionnaire should contain an introductory background which should contain respondents' personal data.

The major types of question or response formats available for researchers include: open-ended questions, multiple choice questions, dichotomous questions (yes or no) and scaling questions.

Interview is a data gathering instrument that enables the seeker of information to have an in-depth knowledge of an issue of concern through face to face interaction with the respondent or informant. Besides interview survey is the calling on respondents on phone in order to obtain information from them through conversation. A panel is a group of individuals who have agreed to provide information to a researcher over a period of time. This means that data can be collected for two or more times. Observation is the technique of watching, monitoring and recording the sequence and details of incidents of particular sorts that a researcher wants to study. This is the only logical means of collecting information.

Experimentation entails deliberate manipulation of one or more conditions (variables) by an investigator in such a way that its impact upon one or more other conditions can be measured. An experiment differs from ordinary observation in that the experimenter or investigator can make changes in the situation so as to see what effect each change has.

Case analysis refers to the intensive analysis of a single or few situations of a phenomenon of the purpose of obtaining information that would lead to a greater understanding of that phenomenon and the possibility of generalization. Case analyses can help the researcher in improving and simplifying the hypotheses to be used in a wide study. This explains why most applications of case analyses are in pilot and explorative studies (Iyiogwe, 2002).

Self-Assessment Exercises 1

1. Research as a critical study is closely tied ….?
2. To test the validity of instruments, research may …?
3. What is the motivating factor behind every research?
4. Research methodology means …..?
5. Mention any five methods of collecting field data.

1.4 **Characteristics of Research**

Features or characteristics of research determine whether it is free of biases, prejudices, and subjective errors or not. These include:

i. Generalized.
ii. Controlled.
iii. Rigorous.
iv. Empirical.
v. Systematic
vi. Reliability.
vii. Validity.
viii. Employs hypothesis
ix. Analytical & Accuracy.
x. Credibility.
xi. Critical

i. Generalized

Researchers often divide the identified population into smaller samples depending on the resource availability at the time of research being conducted. This sample is understood to be the appropriate representative of the identified population therefore the findings should also be applicable to and representative of the entire population. The analytical information obtained from studying these

samples should give a fair idea of total population of being follower of particular ideology, beliefs, social stigmas, driving force, etc.

ii. Controlled

Control implies that, in exploring causality in relation to two variables (factors), you set up your study in a way that minimizes the effects of other factors affecting the relationship. Some variables are classified as controlling factors and the other variables may be classified as possible effects of controlling factors. Laboratory experiments as in pure sciences like chemistry can be controlled but any study that involves societal issues cannot be controlled

iii. Rigorous

Researchers must be careful in ensuring that the procedures followed to find answers to questions are *relevant, appropriate and justified*

iv. Empirical

Empirical nature of research means that the research has been conducted following rigorous scientific methods and procedures. Quantitative research is easier to prove scientifically than qualitative research as the latter often is prone to biases and prejudice.

v. Systematic

Procedures or processes being developed to undertake a study should be carefully drafted to ensure that resources utilization is optimized. Chaotic or disorganized procedures would never yield expected outcomes. The steps should follow a logical sequence to get to the desired outcome.

vi. Reliability

Reliability is defined as the degree to which the result of a measurement, calculation, or specification can be depended on to be accurate. It is the extent to which an experiment, test, measuring procedure, research, research instrument, tool or procedure yields the same results on repeated trials. If any research yields similar results each time it is undertaken with similar population and with similar procedures, it is called to be a reliable research.

vii. Validity & Verifiability

Validity is the extent to which a concept, conclusion or measurement is well-founded and likely corresponds accurately to the real world. The validity of a measurement tool (for example, a test in education) is considered to be the degree to which the tool measures what it claims to measure. Validity is based on the strength of a collection of different types of evidence. In terms of research validity is the strength with which we can make research conclusions, assumptions or propositions true or false. Validation refers to accuracy of measurement whether or not it measures what it is supposed to measure. The findings of a study should be verifiable by the researcher as well as anyone else who wants to conduct the study on similar guidelines/ under similar conditions.

viii. Employs Hypothesis

Every research definitely begins with formulation of a hypothesis. It is a supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation. A hypothesis can be defined as an educated guess about the relationship between two or more variables. Hypothesis may prove to be wrong or null or void after the study is conducted. A hypothesis is an informed and educated prediction or explanation about something. Part of the research process involves testing the hypothesis, and then examining the results of these tests as they relate to both the hypothesis and the world around you.

ix. Analytical & Accurate

Any data collected if does not yield results or is unsuitable to be used for further studies or applications disrupts the purpose of research. Therefore, data collected should be reasonable and free of errors to be easily analysed. Accuracy is also the degree to which each research process, instrument, and tool is related to each other. Accuracy also measures whether research tools have been selected in best possible manner and research procedures suits the research problem or not.

x. Credibility

This has to do with the extent to which an analysis of finding can be treated to be trustworthy. This can only be assured by the use of the best source of information and best procedures in research. Researches based on secondary data are not reliable as such data may have been manipulated or distorted by earlier researchers to suit their work. The research study conducted based on primary data is always reliable and carries more credibility. When researcher gives accurate references in the research the credibility of the research increases but fake references also decrease the credibility of the research.

xi. Criticality

Critical scrutiny of the procedures used and the methods employed is crucial to a research enquiry. The process of investigation must be foolproof and free from drawbacks. The process adopted and the procedures used must be able to withstand critical scrutiny of any sort.

Self-Assessment Exercises 2

1. What do you understand by control variable?
2. The validity of a measurement tool means ….?
3. Why should data collected be reasonable and free of errors?
4. How can we assure credibility in research?

1.5 Purposes/Goals of Research in Hospitality

Generally, research serves many purposes to our daily lives and experiences as stated below.

i. Collect, collate, present and analyse data

ii. Establish reliability and validity of theoretical postulations

iii. Distinguish facts from opinions

iv. Add to existing knowledge

v. Contribute to building of theories

vi. Provides training in problem solving and leadership rules

vii. Provide data for planning and development

viii. Trains scholars for research related jobs.

However, the above are general purpose or importance of research. In hospitality and tourism management which is the thrust of this work, research has certain specific goals. Firstly, marketing research in hospitality and tourism management has to do with a systematic and careful gathering, recording, transmission and analysis of data on issues relating to hospitality and tourism so as to have control over the market forces of demand and supply. Secondly, hospitality and tourism management today is a sensitive, fragile and competitive business which requires appropriate and sound decisions. Thus, firms incur a lot of expenditure on research in order to access reliable data for decision – making. Research therefore provides hospitality, tourism and related firms with the necessary and effective business decision making tools and strategies for survival in the dynamic business world.

Self-Assessment Exercises 3

1. State the major importance of research in hospitality and tourism.
2. Mention the key roles of statistics in research.
3. What is accuracy?

1.6 Summary

This unit discussed the meaning, concepts, characteristics and goals of research. Research is a systematic observation of processes to find better ways to do things and to reduce the effort being put in to achieve an objective and identifying the validity of the targets. It is a subconscious activity that we are involved in at all times. Obviously, the primary purposes of research are discovery, interpretation and documentation for the advancement of human knowledge. It also described the research methodology to consist of the following essential parts: scope of study and area of coverage, sources of data, population of the study, determination of the sample size, sampling procedure/sampling design/sampling plan and administration of research instrument.

1.7 References/Further Readings

Chukwuemeka, E.O. (2002). Research methods and thesis writing: A multi-disciplinary approach. Enugu: Hope- Rising Ventures Publishers.

Eboh, E.C. (2009). Social and Economic Research: Principles and Methods. Enugu: African Institute of Applied Economics.

Iyiogwe, S.O. (2002). Research methodology. New practical approach for report writing in Nigerian universities with vocational statistics. Abakaliki: WillyRose and Applessed Publishing Company.

Gauch, Jr., H.G. (2003). Scientific method in practice. Cambridge, UK: Cambridge University Press.

Groh, Arnold (2018). Research Methods in Indigenous Contexts. New York: Springer.

Ranjit, K. (2014).Research Methodology: A Step by Step Guide for Beginners, Sage Publication.

Rocco, T.S., Hatcher, T & ,.Creswell, J.W. (2011). The handbook of scholarly writing and publishing. San Francisco, CA: John Wiley & Sons.

1.8 Possible Answers to SAEs

Answers to SAEs 1

1. to its objective nature.

2. replicate elements of prior projects or the project as a whole.

3. The existence of a problem and the urge to get it resolved.

4. Research methodology can be defined as a set of specific procedures or techniques used to identify, select, process, and analyze information about a topic.

5. Primary data are called field data and the important methods of collecting primary data are questionnaire, observation, interview, telephone and panel surveys.

Answers to SAEs 2

1. Control implies that, in exploring causality in relation to two variables (factors), you set up your study in a way that minimizes the effects of other factors affecting the relationship.

2. The validity of a measurement tool (for example, a test in education) is considered to be the degree to which the tool measures what it claims to measure.

3. Any data collected if does not yield results or is unsuitable to be used for further studies or applications disrupts the purpose of research.

4. This can only be assured by the use of the best source of information and best procedures in research.

Answers to SAEs 3

1. Research therefore provides hospitality, tourism and related firms with the necessary and effective business decision making tools and strategies for survival in the dynamic business world.

2. To collect, collate, present, analyze data and interpret results.

3. Accuracy is also the degree to which each research process, instrument, and tool is related to each other.

Unit 2: Concepts and Terms Used in Research

Unit Structure

2.1 Introduction

Some basic concepts and terms are frequently used in research methods irrespective of the course of study to denote one thing or another in relation to the laws governing the discipline. Researchers generate concepts by generalizing from particular facts because concepts are based on experiences acquired over time. Concepts can be based on real phenomena and are generalized ideas about something. In scientific research, concepts are the abstract ideas that are being studied. Thus, before conducting any research, the researcher must possess fair idea

of some basic concepts to be used in the research methodology. These concepts and terms are described in this unit.

2.2 Learning Outcomes

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

- Discuss the major concepts used in research; and
- Write short notes on other terms used in research methodology.

2.3 Concepts and Terms Used in Research Methodology

2.3.1 Major Concepts Used in Research

Research as an objective, systematic, controlled and critical activity, planned and directed towards discovery and development of dependable knowledge have major concepts associated with it.

2.3.1.1 Concepts

Every scientific discipline has its own set of concepts which permits communication and research through which researchers can transmit to their colleagues and the entire society, a whole system of experience acquired through research. Thus, concepts are derived from and

developed after series of experiences. Concepts are complex systems of highly abstract ideas, which can be built only by successive experiences in a variety of contexts. They are formally and logically developed ideas about classes of phenomena that a researcher seeks to study; the "building blocks" of theory. Concepts are abstraction from observed events and are generally referred to as universally descriptive words. A person cannot feel or touch concept because it is an abstract. Scientific research tends to be of the explanatory type in that they search for potential explanations of observed natural or social phenomena. Explanations require development of **concepts** or generalizable properties or characteristics associated with objects, events, or people. While objects such as a person, a firm, or a car are not concepts, their specific characteristics or behavior such as a person's attitude toward immigrants, a firm's capacity for innovation, and a car's weight can be viewed as concepts (Bhattacherjee, 2021).

Researchers generate concepts by generalizing from particular facts. The term concept do not refer to a particular thing but to a class of things; rather they help us to connect our ideas. For instance, 'man' as a concept is universal as it can be applied to all human-beings. Concepts as universal terms denote individual things in a special way that have direct empirical reverence. An initial step in empirical research is to define the concepts we are studying. Researchers generate concepts by generalizing from particular facts. Concepts are based on our experiences. Concepts can be based on real phenomena and are a generalized idea of something of meaning. Examples of concepts include common demographic measures: income, age, household size, educational qualification. We can measure concepts through direct and indirect observations.

2.3.1.2 Construct

A construct is an indicator variable that measures a characteristics, or trait. It is an abstract concept that is specifically chosen or created to explain a given phenomenon. A construct may be a simple concept, such as a person's *weight*, or a combination of a set of related concepts such as a person's *communication skill*, which may consist of several underlying concepts such as the person's *vocabulary*, *syntax*, and *spelling*. Constructs are formed from the conglomeration of concept and are created to summarize observations and to provide explanations. They are high level abstractions that cannot easily be described by pointing to a definite object and circumstances. Constructs exist at a higher level of abstraction than concepts and are measured with multiple variables. Complex abstractions such as intelligence, justice, beauty, motivation, happiness, health, trustworthiness are some examples of constructs. Constructs are considered latent variable because they cannot be directly observable or measured. Typical constructs in hospitality and tourism as well as marketing research include brand loyalty, purchase intent, and customer satisfaction. Constructs are the basis of working hypotheses. Constructs used for scientific research must have precise and clear definitions that others can use to understand exactly what it means and what it does not mean. For instance, a seemingly simple construct such as *income* may refer to monthly or annual income, before-tax or after-tax income, and personal or family income, and is therefore neither precise nor clear.

2.3.1.3 Nominal/Conceptual and Operational definitions

In scientific research, nominal or conceptual definition is any attempt to give meaning to a concept. More often than not, a nominal definition is often regarded as neither true nor false, but rather a configuration of empirical characteristics observed, described or postulated and assigned a label. In nominal or conceptual definition, it is wrong to define 'power as the ability to use power'. Such definitions of a circular nature are not particularly useful in scientific research for elaborating the meaning and content of any concept or construct. Conceptual definitions are important in research but should be used consciously. Operational definition is a bridge between conceptual definition and empirical observations and it is specifically related to the real world. Scientific research requires operational definitions that define constructs in terms of how they will be empirically measured. Operational definition has the advantage of making it possible to collect data in terms of observable events. Operational definitions are of two types: measured and experimental operational definitions. In measured operational definition, concepts may be measured. For example, intelligent quotient may be measured in terms of scores obtained in an examination. Experimental definition on the other hand has to do with the production of some condition of experiment by the researcher. For instance, the operational definition of a construct such as temperature must specify whether we plan to measure temperature in Celsius, Fahrenheit, or Kelvin scale. A construct such as income should be defined in terms of whether we are interested in monthly or annual income, before-tax or after-tax income, and personal or family income. Operational definition generally are very essential in any research work because they permit investigators to measure abstract terms in empirical level, which enables their compression (Ominyi and Odo, 2006).

2.3.1.4 Variables

In scientific research, measurement is the assignment of numbers or symbols to phenomena. This requires a scale which provides a range of values or yardstick that corresponds to the presence of the properties of the concept under investigation. A scale provides the rules that associate values on the scale to the concept we are studying. Values derived from measurement can be classified into two broad categories: variables and constants. A variable is a measurable representation of an abstract construct. As abstract entities, constructs are not directly measurable, and hence, we look for proxy measures called variables. Variables are characteristic features, trait or empirical property that can be measured either quantitatively or qualitatively**.** A variable is a quantity that can vary (e.g., from low to high, negative to positive, etc.), in contrast to constants that do not vary (i.e., remain constant). Basically, variables may be classified as independent or dependent variables. Others are moderator, intervening and control variables.

i. **Independent variables** are variables that explain other variables. Stated differently, an independent variable is that which the researcher manipulates or controls in order to ascertain its relationship to observed phenomenon. It is a factor that is measured, manipulated or selected by the experimenter to determine its relationship to an observed phenomenon. For instance, teaching method may be seen as an independent factor in determining level of learning. Here teaching method can be manipulated to bring about changes in level of learning.

ii. Dependent variables are those that are explained by other variables. A dependent variable changes in response to changes in the independent variable or variables. Thus, dependent variable is that variable which the researcher is interested in explaining while the independent variable produces changes in the dependent variable. The dependent variables appears, increases, decreases, or changes as the researcher introduces, removes or changes the independent variable. Any observable effect as a result of this manipulation is the dependent variable. This variable is

the condition (s) which exist later in time than the independent variable. Other variables that can be measured in a research study include intervening, moderating and control variables.

iii. **Intervening or intermediate variables** are those that are explained by independent variables while also explaining dependent variables. These are also intervening variables because they intervene between independent and dependent variables. Factors which theoretically affects the observed phenomenon but cannot be measured, manipulated or seen are said to be mediating variables.

iv. Moderating variables influence the relationship between independent and dependent variables. These factors can be selected or manipulated by the researcher to ascertain if it modifies the relationship of the independent variables to an observed phenomenon. **Control variables** are factors which the researcher must control in order to neutralize or cancel its effect on the observed phenomenon. Control variables are other extraneous variables that are not pertinent to explaining a given dependent variable, but may have some impact on the dependent variable which must be controlled in a scientific research. Control variables could be situational or dispositional. For instance, *noise* could be viewed as a control variable in a comparative study of two teaching methods.

2.3.1.5 Research problem

This is an existing intellectual stimulus that is demanding for a solution in the form of scientific investigation or inquiry. This is tackled by the selection of a research topic, formation of hypothesis and identification of dependent and independent variables.

Self-Assessment Exercises 1

1. We can measure concepts through …?
2. Mention at least five examples of constructs.
3. Operational definitions are of two types …..?
4. How can you tackle a research problem?

2.3.2 Other Terms Used in Research Methodology

2.3.2.1 Research hypothesis

A research hypothesis is defined as a tentative statement which expresses the truth about the relationship existing between variables under study. Hypothesis is not just a guess, but an informed guess. According to Ominyi (2006), hypothesis is an expectation about events based on generalization of the assumed relationship between variables. It is an assumption whose truth or falsity and validity is to be established. Hypothesis is a conjectural statement that could be proved or disproved with the aid of an empirical investigation. It can also be described as a conditional statement that is logically consistent with a theory and can be tested with observations. It is an informed guess or conjecture about any chosen parameter (e.g. mean) of the population. A hypothesis is an educated guess, based on the probability of an outcome. Scientists formulate hypotheses after they understand all the current research on their subject. Hypotheses specify the relationship between at least two variables, and are testable. For a hypothesis to function properly, other researchers must be able to reproduce the results that prove or disprove it (Kara, 2012). A hypothesis is used in an experiment to define the relationship between two variables. The purpose of a hypothesis is to find the answer to a question. A formalized hypothesis will force us to think

about what results we should look for in an experiment. It is a tentative supposition or provisional guess which seems to explain the situation under observation. It is a tentative generalization of the validity of which remains to be tested.

The central objective of hypothesis is to provide a starting point for investigating a research problem in order to achieve the specific objectives without deviating from the purpose of the study. Research hypothesis can be classified into about seven types:

- Simple Hypothesis.
- Complex Hypothesis.
- Empirical Hypothesis.
- Null Hypothesis (Denoted by "H_O")
- Alternative Hypothesis (Denoted by "H_1")
- Logical Hypothesis
- Statistical Hypothesis

However, the two most common types of research hypotheses of interest in this study are null and alternative hypotheses.

a. Null hypothesis denoted by H_o is always stated in negative form and it expresses that there is no significant relationship or difference between two or more variables under study. It is a hypothesis that no difference exists, which implies that the estimate being tested equals a specified value. Thus, any existing differences or relationship found between the estimates is often attributed

to sampling error or chance. A researcher has a null hypothesis when she or he believes, based on theory and existing scientific evidence, that there will not be a relationship between two variables.

b. Alternative hypothesis denoted by H_A or H_1 differs from the null hypothesis because it indicates the existence of statistical difference s or relationships between two or more variables. It is always stated in a declarative form. Alternative hypothesis is often expressed as not equal to, greater or less than the value stipulated under the null hypothesis. Researchers seek to determine whether or not their hypothesis, Ho hypotheses will prove true. Thus, if $H_{o: u = 12, \text{ then}}$ H_A or $H_{1:}$ can be $u \neq 12$, $u > 12$ or $u < 12$.

c. Characteristics & Qualities of a Good Hypothesis

A working hypothesis is a hypothesis that is provisionally accepted as a basis for further research in the hope that a tenable theory will be produced, even if the hypothesis ultimately fails.

- Power of Prediction. One of the valuable attribute of a good hypothesis is to predict for future.
- Closest to observable things. A hypothesis must have close contact with observable things.
- Simplicity.
- Clarity.
- Testability.
- Relevant to Problem.

Whenever a hypothesis is formulated, the most important thing is to be precise about what one's variables are, what is the nature of the relationship between them might be, and how one can go

about conducting a study on them by way of hypothesis testing. Hypothesis testing involves the following steps:

i. Clearly formulate and state the hypotheses (null and alternative).

ii. Choose the method of statistical test to use (the standard error test, the t-test and the F-test).

iii. Choose the level of significance to use in testing the formulated hypothesis

iv. Calculate the value of the chosen statistical test to test the hypothesis

v. State decision rule to decide the acceptance or rejection of the null hypothesis.

vi. Compare the calculated value with the theoretical or table value.

vii. Make conclusion on the acceptance or rejection of the null hypothesis.

d. Functions of Hypotheses

A **hypothesis** is either a suggested explanation for an observable phenomenon, or a reasoned prediction of a possible causal correlation among multiple phenomena and it performs the following functions:

i. Development of Research Techniques

ii. Separating Relevant From Irrelevant Observation

iii. Selecting Required Facts

iv. Direction of Research

v. Acts as a Guide

vi. Prevents Blind Research

vii. Accuracy & Precision

viii. Link between Theory & Investigation

ix. Link between Assumption & Observation

x. Provide answer for a Question

xi. Save Time, Money & Energy.

2.3.2.2 **Research theory**

If enough evidence accumulates to support a **hypothesis**, it moves to the next step known as a **theory.** Scientific knowledge is represented as a collection of theories derived using the scientific method. A **theory** is a tested, well-substantiated, unifying explanation for a set of verified, proven factors in the scientific method and becomes accepted as a valid explanation of a phenomenon. Research theories are statements of beliefs but when theories are tested and proved they become laws. Theory is a set of logically consistent ideas about the relationships between empirical phenomena or concepts that permits those ideas to be tested using observations. A scientific theory is defined as a system of constructs (concepts) and propositions (relationships between those constructs) that collectively presents a logical, systematic, and coherent explanation of a phenomenon of interest within some assumptions and boundary conditions (Rogers, 2005). Theories explain certain events or phenomenon which provide basis for identifying research problems.Theories are usually used to help design a research question, guide the selection of

relevant data, interpret the data, and propose explanations of the underlying causes or influences of observed phenomena. Theories should explain why things happen, rather than just describe or predict. A theory can be seen as a verifiable fact but a fact refers to an empirically verifiable observation about phenomena which helps to initiate a theory. Thus, theory plays an active role in uncovering facts, while facts themselves equally play significant role in developing theories which scientific research greatly depends upon.

Research theory equips the researcher with a way of looking at reality. It is a systematically organized knowledge of varying degrees of generalization with a view to the eventual specification of relationships among empirical tests. Test of a research theory is the degree to which its formulation seem congruent with our own perception of the real world situation (Chukwuemeka, 2002). Theory is a deductive concept which accommodates common properties belonging to more than one events, issues, cases or variables. A good and pertinent research theory places a study at advantages since it serves as a guide and provides a rationale basis for interpreting research results. A scientific theory is a system of constructs (concepts) and propositions (relationships between those constructs) that collectively presents a logical, systematic, and coherent explanation of a phenomenon of interest within some assumptions and boundary conditions. Thus, a theory is a set of logically consistent ideas about the relationships between empirical phenomena (concepts) that permits those ideas to be tested using observations. Because theories are also intended to serve as generalized explanations for patterns of events, behaviors, or phenomena, theoretical explanations are generally nomothetic in nature. Theory is not data, facts, typologies, taxonomies, or empirical findings. A collection of facts is not a theory, just as a pile of stones is not a house.

The use of theories in research has many benefits. Theories provide the underlying logic of the occurrence of natural or social phenomenon by explaining what are the key drivers and key outcomes of the target phenomenon and why, and what underlying processes are responsible driving that phenomenon. Again, theories aid in sense-making by helping us synthesize prior empirical findings within a theoretical framework and reconcile contradictory findings by discovering contingent factors influencing the relationship between two constructs in different studies. Thirdly, theories provide guidance for future research by helping to identify constructs and relationships that are worthy of further research. Fourthly, theories can contribute to cumulative knowledge building by bridging gaps between other theories and by causing existing theories to be reevaluated in a new light. However, theories can also have their own share of limitations. As simplified explanations of reality, theories may not always provide adequate explanations of the phenomenon of interest based on a limited set of constructs and relationships. Theories are designed to be simple and parsimonious explanations, while reality may be significantly more complex. Furthermore, theories may impose blinders or limit researchers' "range of vision," causing them to miss out on important concepts that are not defined by the theory. Examples of research theories include:

i. Agency Theory (also called principal-agent theory), a classic theory in the organizational economics literature, was originally proposed by Ross (1973);

ii. Theory of Planned Behavior which was postulated by Azjen in 1991;

iii. Innovation Diffusion Theory which concept was first studied by French sociologist Gabriel Tarde, but the theory was developed by Everett Rogers in 1962.

Self-Assessment Exercises 2

1. A hypothesis is used in an experiment to define …?
2. The two most common types of research hypotheses are…?
3. Outline the three examples of research theories.
4. State the role of facts in research.

2.4 Summary

This unit discussed the concepts and terms used in research methodology such as hypotheses and research theory. In its most elementary stage the hypothesis may be a guess, imaginative idea which becomes the basis for further investigation while research theories explain why things happen, rather than just describe or predict as examined herein.

2.5 References/Further Readings

Chukwuemeka, E.O. (2002). Research methods and thesis writing: A multi-disciplinary approach. Enugu: Hope- Rising Ventures Publishers.

Eboh, E.C. (2009). Social and Economic Research: Principles and Methods. Enugu: African Institute of Applied Economics.

Iyiogwe, S.O. (2002). Research methodology. New practical approach for report writing in Nigerian universities with vocational statistics. Abakaliki: WillyRose and Applessed Publishing Company.

Gauch, Jr., H.G. (2003). Scientific method in practice. Cambridge, UK: Cambridge University Press.

Groh, Arnold (2018). Research Methods in Indigenous Contexts. New York: Springer.

Ominyi, C.N. and Odo, F.A. (2006). Research and statistics in education, management and social sciences. Enugu: Mason Publishers.

Ranjit, K. (2014).Research Methodology: A Step by Step Guide for Beginners, Sage Publication.

Rocco, T.S., Hatcher, T & ,.Creswell, J.W. (2011). The handbook of scholarly writing and publishing. San Francisco, CA: John Wiley & Sons.

Rogers, E. (2005). Diffusion of Innovations. New York: The Free Press.

2.9 Possible Answers to SAEs

Answers to SAEs 1

1. We can measure concepts through direct and indirect observations.

2. Complex abstractions such as intelligence, justice, beauty, motivation, happiness, health, trustworthiness are some examples of constructs.

3. Operational definitions are of two types: measured and experimental operational definitions.

4. A research problem is tackled by the selection of a research topic, formation of hypothesis and identification of dependent and independent variables.

Answers to SAEs 2

1. A hypothesis is used in an experiment to define the relationship between two variables.

2. The two most common types of research hypotheses are null and alternative hypotheses.

3. Examples of research theories include:

i. Agency Theory (also called principal-agent theory), a classic theory in the organizational economics literature, was originally proposed by Ross (1973);

ii. Theory of Planned Behavior which was postulated by Azjen in 1991;

iii. Innovation Diffusion Theory which the concept was first studied by French sociologist Gabriel Tarde, but the theory was developed by Everett Rogers in 1962.

4. Facts play significant role in developing theories which scientific research greatly depends upon.

Unit 3 Types and Classification of Research

Unit Structure

3.1 Introduction

3.1 Introduction

There seems to be no unanimous agreement among researchers on the system of classification of research". The fact that practically every research book suggests a different system of classification provides convincing evidence that there is no generally accepted scheme. This is because most of the types or classifications of research are intertwined or overlapping. Research can be classified into various categories depending on the perspective under which the research activity is initiated and conducted leading to the existence of different typologies of research in line with the various schemas reflecting various perspectives of classification. While there is no exclusive universal hierarchical classification of research in social sciences, there exist functional typologies by which various social science disciplines (e.g. political science, sociology, economics, etc.) derive their tailored classifications as shown in this unit. In this section, we have discuss the more common

classifications of research in the literature of hospitality and tourism management. However, in this context, we shall classify research into three: by purpose, method and evidences required or method of data analysis.

3.2 Learning Outcomes

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

- Discuss the classification of research based on purpose;
- Analyse research classification based on method and evidence required;
- Describe the classification of research on the basis of method of data analysis.

3.3Types and Classification of Research

3.3.1 Classification Based on Purpose

This classification falls into five basic categories:

i. Basic or pure research

ii. Applied research.

iii. Action research

iv. Evaluation research

v. Research and Development (R & D)

i. **Basic or Pure Research**

This is a fundamental research undertaken to discover basic truths or principles. It is the type of research that is not concerned with the usefulness of the findings to practical situations but interested in the building of theories. Basic or pure research is concerned with the production of result and finding which lead to development of theory. This is also called academic research which is considered to advance scientific knowledge through the verification of hypothesis used in testing theories. . It refers to research aimed at testing conceptual hypothesis or developing theoretical principles. Pure research entails doing research to gather knowledge for knowledge's sake. Basic research is used to depict studies directed at finding information possessing a broad base of application and adds to the existing stock of scientific knowledge.

ii. Applied Research

This is also called field research and mostly oriented towards finding solutions to specific behavioural and organizational problems of the society with the aim of ameliorating the situation. It is conducted for the purpose of applying or testing theory and evaluate its usefulness in solving problems. Applied research is concerned with the usefulness of ideas or theories on practical situations. It does not give birth to theory formulation but seeks to establish the workability of a given theory to practical situations. It is data-based and gives conclusions that are verifiable through observation or experimentation. Empirical research is appropriate where a study seeks to prove if certain variables affect others. While pure research and applied research may differ in approach, there is interplay between them. Good theoretical research can be applicable to practical Problems and practical/applied research

can enable advancements in theory. Thus, both basic and applied research are useful in their contributions to the development of knowledge.

iii. **Action Research**

As the name suggests, action research enables the researcher to take immediate action based on the outcome of the research. Action research is undertaken specifically for the purpose of resolving a particular problem, aiming at collection of information which can be immediately used to help make practical decisions relating to development planning, programme design and implementation. Such research concerns the diagnosis of a situation. Action research is applied in schools to solve specific classroom problems. For instance, it can be applied to the problem of stealing in the classroom. It is usually taken by a teacher or group of teacher in school to solving specific problem so that the result can be generalized. By attempting to answer the question of 'what is', action research seeks to describe events and situations as they appear in the society.

iv. **Evaluation Research**

Evaluation research is a systematic process of collecting empirical data for analysis, on the basis of which decisions could be made about individuals, programmes, methods and events. It is concerned with making decisions on the effectiveness of a given system to bring about timely modifications in the programme or system where necessary so as to entrench efficiency in the system.

v. Research and Development (R and D)

Research and development programmes aims are at developing and testing more efficacious and other useful products such as textbooks, laboratory, teaching curricula etc. It involves the testing and revision of academic and household materials to ensure that the required

standard is met. R and D is aimed at developing and testing all products to ensure their effectiveness. Although costly, the essence of research and development is to establish pattern of growth in a country's sojourn in some place or situation and this is very useful in providing valuable means of achieving educational and overall improvement in the society. R and D studies could be undertaken in cross-sectional or longitudinal manner as the case may be.

Self-Assessment Exercises 1

1. Another name for basic or pure research is …?
2. Action research enables the researcher to …?
3. R and D is aimed at …..?
4. Applied research is concerned with…?

3.3.2 Classification Based on Method and Evidence Required

Classification of research based on the method employed can also be categorized into five as well. These are:

i.Historical Research

ii.Descriptive (Survey) Research

iii.Correlational Research

iv. Experimental Research

v.Casual-comparative (ex-post-facto) Research.

i. Historical Research

Historical research is an investigation into the past events with a view to exposing and understanding the past, and relating the result to the present so as to have a sound base for predicting the future. It refers to past records and relies on the analysis of past events to determine the present and project into the future. The objective of historical research is to determine what had been in existence. This provides us with the truth about the past which could be used in the interpretation of current data that can link the past to present and possibly project into the future. This type of research is based on oral evidence and other useful documents such as diaries, case history, autobiography, logbook, books, journals, magazines etc.

ii. **Descriptive or Survey Research**

Survey and experimental research methods are approaches to descriptive research. This is concerned with the collection, collation, presentation, analysis and interpretation of obtained data for the purpose of describing practical beliefs and attitudes of objects, individuals, groups and institutions in order to list, describe, explain, classify, compare, contrast, analyse and give appropriate interpretation to the situation or events. As the name implies, it is intended to describe the present status of an issue or a problem which is analyzed based on the available data and so does not require hypothesis to begin with. For instance, if a guest in a tourist centre complains about a faulty shower in the bathroom out of ignorance or inability to use it, it has to be resolved delicately and not by pointing out to him that he is not aware of new technology.

A descriptive study entails the systematic collection and presentation of data to give a clear picture of a particular situation. Descriptive research is fact-finding in nature. It focuses on selective

dimensions of a phenomenon and measures them in a systematic and precise manner. All over the literature on research methods, descriptive studies employ a number of methods including case studies, sample survey, census and aerial survey. Descriptive or survey research is based on information, interviews, (oral, written, structured, unstructured), inventories, rating scales, self-report, observations etc.

iii. Correlational Research

Correlational research is used to establish the magnitude and direction of relationship between variables under study. The objective of this research is to measure the degrees of relationship between two or more variables (dependent and independent) for the purposes of making predictions about their relationship. This seeks to discover if two variables are associated or related in some way, using statistical analysis, while observing the variable. E.g. If the heat is reduced or increased during cooking how does the food react to it.

iv. **Experimental Research**

Experimental research is associated with scientific investigation in the field of natural, physical, biological, medical, engineering and agricultural sciences which involves laboratory experimentation. Experimental research involves conducting experiment for research purpose. This is a method used to establish a cause and effect relationship between two variables or among a group of variables. The independent variable is manipulated to observe the effect on the depended variable. E.g. The change in response to between groups of foreigners treated to welcome drinks and freshener tissues and the one that is simply welcomed and allocated rooms in a hurry due to peak hours of check in and check out (Kara, 2012). Experimental research refers to studies whereby independent variable (the intervention) is deliberately manipulated by the researcher

under known, clearly defined and controlled conditions. In an experimental design, a researcher allocates randomly selected individuals to at least two groups. One group known as the experimental group is subjected to the intervention or treatment while known as the control group is not exposed to the intervention. Both experimental and groups should be investigated under conditions that are identical, except for the exposure of intervention applied to the experimental group. This is absolutely necessary to minimize variation between them.

The classical experimental study has three integral requirements; manipulation or treatment or intervention, control and randomization. Manipulation means that the researcher does something to one group of subjects in the study. Control implies that the researcher introduces control groups to compare with the experimental group. Randomization means that the researcher ensures that each subject is given an equal chance of being assigned to either group. By this, the researcher endeavors to hold all other factors constant so that differences between the control and experimental groups can be linked to the treatment/intervention being applied. The experimental research involves investigating possible cause and effect relationship by exposing one or more experimental groups to a treatment and one or more control groups not receiving the treatment.

v. **Causal-Comparative (Ex-Post-Facto) Research**

Ex-post literally means after the fact or retrospectively**.** Causal-comparative or ex-post-facto research involves determining the effect of past factor on what happens thereafter. This type of research is an investigation to examine the extent a factor to which a group is exposed to in the past will influence the same group, different from other groups similar to the aforementioned group but were not exposed to the said factor at the time of the study. It is a situation where some effects

are attributed to some cause without attempt to manipulate the independent variable because they cannot be controlled in the study area. Ex-post facto is a systematic empirical inquiry in which the researcher has no direct control of independent variables because their manifestations have already occurred inherently and cannot be not manipulated. An example here is the influence of the length of the leg in long distance races. Here, researchers ask themselves, what factors seem to be associated with certain occurrence conditions or aspects of behavior. Ex-post facto is a method of testing out possible antecedents (past) of events that have happened (Chukwuemeka, 2002).

3.3.3 Classification based on method of data analysis

There are three major types of research based on method of data analysis. These are:

i. Qualitative or Non-computational Research

ii. Quantitative or Computational Research

iii. Triangulation of multiple perspectives.

i. Qualitative or Non-Computational Research

Qualitative research applies to designs, techniques and measures whose data are inform of words rather than numbers which are often grouped into categories. These include non-statistical data approach which involves direct observation, participant observation and interview method of man-made or natural events. A qualitative research design is used to explore the meaning and understanding of complex social environments, like the nature of experiences gained by a tourist by reading about the texts and stories shared by them. It also intends to understand, describe or discover the findings. The researcher is usually the primary instrument that formulates

the question and interprets the meaning of a data. The data used are mostly documented words from interview, newspapers videos etc. More than one type of data is collected during this research, from the field, where the participants are. In other words, the research goes beyond the intended scope, so making it emergent because the method of research changes and different types of data might be collected as the research goes on.

ii. Quantitative or Computational Research

This is associated with the traditional mode of scientific research that moves from theory to operationalization and to observation and then the verification of logically deduced theories. This involves designs, techniques and measures that produce discrete numerical and quantifiable data which can be presented and coded. Vigorous methods of data collection and analysis are applicable here. Its emphasis is usually on statistical data and the utilization of same for hypothesis testing.

iii. Triangulation of Multiple Perspectives

Triangulation is the process of using two or more methods to verify the validity of the information being collected. According to Eboh (2009), the logic of triangulation is to use findings from one method, say qualitative investigation to check against a quantitative study, in a complementary manner. It can take the form of incorporating different points of view as active participants in the research team (women/men, social scientists/natural scientists or technical specialists, youth/elders, insiders/outsiders, etc). It often involves eliciting a wide range of opinions, attitudes and behaviour from interviewees (women/men, old/young, diverse ethnic groups, different socio-economic groups/strata, etc.). It can also mean using a variety of information gathering methods

by addressing the same issue using several different tools (historical interviews, spatial maps, local calendars, etc.).

Self-Assessment Exercises 2

1. State the objective of correlational research.
2. Ex-post facto is a method of testing …?
3. What is triangulation?
4. Mention the emphasis of a qualitative research.

3.4 Summary

Research can be classified into various categories depending on the perspective under which the research activity is initiated and conducted. However, the basic classification of research is based on the purpose of the research. In this unit, three basic classifications were made. First classification is based on purpose: basic or pure research, applied research, action research, evaluation research and development. The second classification is based on method and evidence required: historical research, descriptive (survey) research, correlational research, experimental research and casual-comparative (ex-post-facto) research while the third classification is based on method of data analysis such as: qualitative or non-computational research, quantitative or computational research, and triangulation of multiple perspectives.

3.5 Glossary

Concepts: are abstraction from observed events and are generally referred to as universally descriptive words.

Fact: refers to an empirically verifiable observation about phenomena which helps to initiate a theory.

Hypothesis: is a tentative supposition or provisional guess which seems to explain the situation under observation.

Questionnaire: is a list of questions or statements which require respondents to make responses. It is a data gathering instrument in which respondents are given standard or uniformed questions.

Reliability: is the extent to which an experiment, test, measuring procedure, research, research instrument, tool or procedure yields the same results on repeated trials.

Theory: is a verifiable fact

Validation: refers to accuracy of measurement whether or not it measures what it is supposed to measure.

3.6 References/Further Readings

Advanced Search Appearance Support (2020). Definition of research. Western Syndney University. www.westernsyndney.edu.au. Accessed on 11/8/2022.

Bacharach, S. B. (1989). "Organizational Theories: Some Criteria for Evaluation," Academy of Management Review (14:4), 496-515.

Bhattacherjee, A. (2021) Concepts, Constructs, and Variables. **University of South Florida. Accessed on 11/8/22 from:** info@libretexts.org.

Chukwuemeka, E.O. (2002). Researh Methods and Thesis Writing: A multi-disciplinary approach. Enugu: Hope-Rising Ventures Publishers.

Eboh, E. C. (2009). Social and Economic Research: Principles and Methods. African Institute for Applied Economics, Enugu.

Kara, O. (2012). Researching for Hospitality and Tourism Management. An unpublished course material of the Uttarakhand Open University.

Ominyi, C.N. and Odo, F.A. (2006). Research and statistics in education, management and social sciences. Enugu: Mason Publishers.

Rogers, E. (2005). Diffusion of Innovations. New York: The Free Press.

3.7 Possible Answers to SAEs

Answers to SAEs 1

1. This is also called academic research.

2. As the name suggests, action research enables the researcher to take immediate action based on the outcome of the research.

3. R and D is aimed at developing and testing all products to ensure their effectiveness.

4. Applied research is concerned with the usefulness of ideas or theories on practical situations.

Answers to SAEs 2

1. The objective of this research is to measure the degrees of relationship between two or more variables (dependent and independent)
2. Ex-post facto is a method of testing out possible antecedents (past) of events that have happened.
3. Triangulation is the process of using two or more methods to verify the validity of the information being collected.
4. Its emphasis is usually on statistical data and the utilization of same for hypothesis testing.

MODULE 2 RESEARCH DESIGN

Unit 1: Meaning and Relevance of Research Design

Unit Structure

1.1 Introduction

Research design can be seen as a plan of investigation which specifies the sources and types of information relevant to the research question. It specifies the blueprint or strategy to be employed in data collection, analysis and interpretation of research results. Social and economic research have two broad approaches: experimental and non-experimental research designs. In seeking to explain and understand human behaviour, experimental and non-experimental social research differs in the degree of control they have over the variables. While experimental research manipulates variables suspected to produce effect, non-experimental research merely observes the variables and their effects. Definitional variants of the dichotomy between qualitative and quantitative research can be linked to varying criteria. They include the nature of the subject under investigation, the type of data required, the method of collecting data, contextuality and the nature of the outcome of the research. Research design ensures that the results of a study are as accurate and unbiased as possible.

1.2 Learning Outcomes

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

- Discuss research design in various ways;
- Analyse the properties of a good research design; and
- State the relevance of research design.

1.3 Meaning and Relevance of Research Design

Research Design is important as it guides the researcher to identify the correct methods of data collection and analysis, conditions in which the activity of research shall be carried out and estimation of the funds to be utilized for it; maintaining its connectivity to the purpose of research. It seeks to answer the questions what and how. It is a process to identify or develop a hypothesis that is further tested using other techniques. A qualitative research design is used to explore the meaning and understanding of complex social environments, like the nature of experiences gained by a tourist by reading about the texts and stories shared by them. It also intends to understand, describe or discover the findings. The researcher is usually the primary instrument that formulates the question and interprets the meaning of data.

1.3.1 Meaning of research Design

Research design simply means drawing an outline from which research is executed. This has to do with a number of logical decisions to be taken before the commencement of a research project. It

is the conceptual structure within which research would be conducted. A research design is the operational paradigm and conceptual frame of the scientific enquiry. It gives shape, form and identity to the research activity. Research design serves as the anchor of the scientific study. It provides smooth sailing and enables the evaluation of the scientific exercise. It is the research plan or blueprint of action. According to Eboh (2009), research design is the deliberate strategic approach used in conducting a scientific enquiry. It answers the questions about a scientific inquiry in terms of what, where, when, how and by what means? The principles of research design involve a set of rules and norms that guide the choice and development of a research design. The choice of research design involves decisions relating to: what entities (individuals, groups, communities, organisations) should be studied? What aspects of characteristics of these entities are of interest? Where will the study be located? Why is the study being undertaken? What periods of time will the study cover? Where can data be found? What type of data is needed? And what kinds of relationships will be anticipated/studied?

Thus, a research design could be seen as a comprehensive plan for data collection whose purpose is to answer research questions and test research hypotheses. It is the plan, structure, and strategy of investigation, conceived as to obtain appropriate responses to research questions and control variance as well. The function of research design is to provide for the collection of relevant information with smallest possible input in terms of effort, time and money. A research is the specification of methods and procedures for acquiring the information needed. It is the overall operational pattern or framework of the project that stipulates what information is to be collected from which sources by what procedures. Broadly speaking, there are two types of research designs, these are descriptive (survey) and experimental designs. It is noteworthy that research design and research methodology are interrelated and these have been discussed extensively in unit two of

module 1 as types and classification of research. The preparation of research design, appropriate for a particular research problem, involves the consideration of the following:

- Objectives of the research study.
- Methods of Data Collection to be adopted
- Sources of information—Sample Design
- Tools for Data collection
- Data Analysis Tools: qualitative and quantitative (Brent & Goeldner, 2014).

A good research design is characterized by its flexibility, effectiveness and suitability. A properly developed *research design* is the one that results in minimal or no error at all if everything goes as planned for. It is important to have clarity of the research question for the objectives to be achieved. Therefore researcher may have to create mix of various design approaches to create a suitable one for the problem being addressed. It facilitates the smooth sailing of the various research operations, thereby making research as efficient as possible yielding maximal information with minimal expenditure of effort, time and money.

1.3.1.1 Steps in developing a sound research design

Steps in developing a sound research design must include:

- Classify the intended outcome i.e. what needs to be understood.
- Develop the research question.
- Understand what needs to be measured.

- Select the population as per the study taken up.
- Identify the ideal data collection method.
- Construct interconnected characteristics.
- Use correct analysis tools.
- Decide how the findings of the study shall be published.

Self-Assessment Exercises 1

1. Research design simply means …?
2. The function of research design is to provide …?
3. Mention the two types of research design.
4. A good research design is characterized by its flexibility, effectiveness and…?
5. A properly developed *research design* is the one that results in minimal or …?

1.3.2 **Properties of good Research Design**

Kara (2012) outlined the characteristics or properties of a good research design to include the following:

- **Objectivity:** Objective findings may be achieved by allowing more than one person to

agree between the final scores/ conclusion of the research.

- **Reliability:** Researcher should ensure that research questions are framed judiciously to make it reliable and provide similar outcomes. Thus the results obtained should be similar if the research is conducted in identical conditions and is repeated time and again.

- **Generalization:** The information collected from given sample must be utilized for providing a general application to the large group of which the sample is drawn.

- **Ethical:** It should be acceptable and be free of practices or procedures that may not be honest or may give error /bias.

- It should be proficient in obtaining the most reliable and valid data;

- A good research design should be able to address any situation wherein any unexpected events can be accommodated. ;

- It also helps a researcher arriving at flawed / misunderstood conclusions;

- It can adequately control the various threats of validity, both internal and external.

1.4 **Relevance of Research Design**

Research designs are aimed at achieving the following objectives.

- It helps to gain familiarity with a phenomenon.
- It helps the researcher to achieve new insights into the problem.
- It helps to formulate a more precise research problem and hypotheses
- It reduces inaccuracy.
- It helps to get maximum efficiency and reliability.

- It eliminates bias and marginal errors.

- Research design stands for advance planning of the method to be adopted for collecting the relevant data and the techniques to be used in their analysis, keeping in view the objective of the research and the availability of staff, time and money.

- It minimizes wastage of time.

- It is helpful for collecting research materials.

- It is helpful for testing of hypothesis.

- It gives an idea regarding the type of resources required in terms of money, man-power, time, and efforts.

- It Provides an overview to other experts.

- It Guides the research in the right direction.

Self-Assessment Exercises 2

1. Define reliability of a research result.
2. Ethical standard means ----?
3. Another name for descriptive design is ----?

1.5 Summary

A research design is the operational paradigm and conceptual frame of the scientific enquiry that gives shape, form and identity to the research activity. It is the anchor of the scientific study which provides smooth sailing and enables the evaluation of the scientific exercise. All *good research designs are* characterized by its flexibility, effectiveness and suitability. These help to reduce inaccuracy in data collection, analysis and testing of research results among other benefits as discussed in this unit.

1.6 References/Further Readings

Brent R.J. and Goeldner, C.R. (2014). Travel, Tourism, and Hospitality Research: A Handbook for Managers and Researchers. UK: Wiley Publishers Limited.

Eboh, E.C. (2009). Social and economic research: Principles and methods (2nd Edition). Enugu: African Institute, for Applied Economics.

Kara, O. (2012). Researching for Hospitality and Tourism Management. An unpublished course material of the Uttarakhand Open University.

Kothari C.R. (2011). Research Methodology. UK: New Age International.

1.7 Possible Answers to SAEs

Answers to SAEs 1

1. Research design simply means drawing an outline from which research is executed.

2. The function of research design is to provide for the collection of relevant information with smallest possible input in terms of effort, time and money

3. There are two types of research designs, these are descriptive (survey) and experimental designs

4. A good research design is characterized by its flexibility, effectiveness and suitability

5. A properly developed *research design* is the one that results in minimal or no error at all if everything goes as planned for.

Answers to SAEs 2

1. Reliability means that the results obtained should be similar if the research is conducted in identical conditions and is repeated time and again

2. It should be acceptable and be free of practices or procedures that may not be honest or may give error /bias.

3. Survey.

Unit 2: SAMPLING

Unit Structure

2.1 Introduction

In every study, the researcher goes about in search of information relevant to the study. His or her source of information will depend on the type and scope of the research. Research data are collected from both primary and secondary sources. Primary data are obtained with the aid of questionnaire, personal observation and interview while secondary data are gathered from newspapers, journals books and other secondhand information sources. The researcher is expected to describe every bit of his/her activities as it affects every member of the research frame not minding the population size. The demand of research especially when large population size is involved brought about the idea of sample and sampling as a way forward.

2.2 Learning Outcomes

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

- Distinguish between population and sample
- Analyse the concept of sampling.

2.3 Sampling

2.3.1 Population and Sample Size

i. Population refers to a group of people or objects existing within a specified geographical area. An example is the population of Kaduna state, Nigeria. In research, population defines the limit which the research finding is applicable and described in such a manner that the result of the investigation can be generalized unto its members. Population can be classified into finite and infinite. Finite population has definite or countable number of elements while infinite population has indefinite number of elements that are uncountable. It is therefore evident that research population is of various sizes and every researcher has the duty to clearly define and demarcate his or her research population in each case. This makes sample selection from a given population very imperative.

ii. Sample or sample size is a smaller group of elements selected through a definite procedure from a given population. It is a representative group selected from the entire population for the purpose of research with a view to making generalisations about the whole population based on the research result. In drawing a good sample, attempt is often made to ensure that most if not all the population characteristics are present in the selected sample. This will determine the closeness of the estimates of the parameters to the validity of conclusion and generalisations made about the

population. The failure of any research sample to represent the population from which it was drawn results to sampling error.

iii. Sampling error is the difference between the result obtained from a sample and that which could have been obtained if the whole population is used for the research study. This could also occur when the complete survey of the population is not carried out before taking a sample size for estimating the characteristics of the population. The smaller the sample error, the greater the precision of the estimate and the more acceptable the research result. Thus, researchers are advised to adopt sampling techniques that will guarantee a high degree of representation in the selected sample size to enhance the validity of their research results. However, the study of the entire population could be possible in certain research studies under the following conditions:

- **w**hen the population of study is very small;
- time available for the research is small;
- material and human resources are inadequate;
- when complete population count is the primary objective of the study.

Besides, the use of the population sample for a study is necessary when:

- the population size is very large;
- time, human and material resources available are adequate.

Self-Assessment Exercises 1

1. Population can be classified into --- and ----?
2. The smaller the sample error, the greater the ----?
3. Mention any 3 conditions that can allow a researcher to study the entire population.
4. State the two conditions that necessitate the use of sample in research study.

2.3.2 Concept of Sampling

Sampling is the selection of the part of the population on the assumption that the part is the representative of the whole. Sampling is a process used in statistical analysis in which a predetermined number of observations are taken from a larger population. Sampling is the selection of a number of study units from a predefined study population or universe. All items in any field of inquiry make up a population or universe. If all the items in the population are enumerated, the process is known as census. A sample is intended to be a representative or microcosm of the population of study. It is the act, process, or technique of selecting a representative part of a population for the purpose of determining parameters or characteristics of the whole population. In a study involving a very large population, the part of the population from which information is collected is called a sample. A sample is justified based on cost and accuracy grounds. It is cheaper, faster and can provide sufficiently accurate results. Sampling permits economy of time, financial and scientific resources.

Sampling is an essential key procedure of the scientific method, not just in hospitality and tourism management but in all areas of human endeavor. Sampling is well advanced in biological and agricultural sciences as well as in physics and chemistry, where every experiment is merely a sample of all possible observations which could be made. All experiments are indeed samples from a larger universe of all possible experimental situations (Eboh, 2009). A sample design includes the modules, techniques or procedures to be used in identifying the items for the sample. Sample

design also defines the number of items to be included in the sample, that is the size of the sample. Sample design is determined before data is collected. The choice of determining sampling design depends on the individual and the type of research undertaken.

i. Characteristics of a good sample design

According to Kothari (2011), a good sampling design should possess the following characteristics:

- **Proportionality**: Sample design must result in a truly representative sample. This means that the sample selected should be exactly or almost similar to the population it represents in terms of data and characteristics.

- **Error Free:** Sample design should reduce the probability of errors. The minimum numbers of errors in any sample ensure correct data obtained and analyzed.

- **Budgeted**: Sample design must be practical and be within the limits of funds available for the research study.

- **Unbiasedness:** Sample design should be able to control systematic bias.

- **Generalization**: Sample should be such that the results of the sample study can be applied, in general, for the universe with a reasonable level of confidence.

ii. Sampling frame is a list of every member (or unit) of the population from which the sample will be drawn. It denotes the set of all cases from which the sample is actually selected. In constructing the sampling frame, the definition of the unit of analysis is followed by the definition of the population of interest, that is, the collection of units that constitute the population. Defining the population requires identifying the target population about which the results are to be generalised and constructing the sampling frame. In defining the target population, care should be

taken to precisely specify the geographical and time dimensions, and rules of membership. This means demarcating the geographical boundaries as well as time frame. The study population has to be clearly defined, for example, according to sex, age, residence or it may consist of villages, institutions, development projects. It is important to define what constitutes a member of the population in an unambiguous manner and language. The clear definition of the target population lays the basis for the operationalization of the definition, through the construction of the sample frame. The accuracy of any sampling procedure depends largely on the accuracy of the frame. Incidentally, sampling frame is often riddled with problems. The problems commonly associated with sampling frames include: non-coverage or omission, duplication, blank or foreign inclusions, and incidence of clustering.

iii. The Requirements of Good Sampling

There are two basic requirements in sampling: a sample must be representative and it must be adequate (of sufficient size).

- **The Representativeness of Samples**

To be representative, a sample should provide a close approximation of the characteristics of the target population from which it is drawn. A sample ought to be a "true" cross section of the whole population of study. A representative sample allows for valid conclusions for the whole study population. It will be wrong to make conclusions about a sub-section of a population unless it truly represents the population. Here, a researcher will be wise to include those study units that are extremes in the study population, with respect to a certain characteristic.

- **The Adequacy of Samples**

A sample is adequate when it is of sufficient size to allow confidence in the stability of its characteristics. In deciding the sample size, the researcher should be guided by considerations whether the sample size is small enough to be affordable and manageable and yet large enough to give reasonably reliable estimates of the population being measured. Hence, the actual sample size is usually a compromise between what can be afforded and the scope of data requirements. As a rule of thumb, if the basic requirement of representativeness has been met, we are more likely to get a better picture of the population, the larger the sample.

Self-Assessment Exercises 3

1. Sampling is the selection of ….?
2. The choice of determining sampling design depends on … and …?
3. Sampling frame means ….?
4. Mention the two basic requirements of good sampling.

2.6 Summary

This unit discussed the concepts of population and sample. We observed that rather than spend long periods doing the slow and tedious work of analyzing a mass of material from one point of view, the researcher is able, by drawing a sample, to examine a smaller amount of materials from many points of view that is, intensive analysis of fewer cases.

2.7 References/Further Readings

Brent R.J. and Goeldner, C.R. (2014). Travel, Tourism, and Hospitality Research: A Handbook for Managers and Researchers. UK: Wiley Publishers Limited.

Eboh, E.C. (2009). Social and economic research: Principles and methods (2nd Edition). Enugu: African Institute, for Applied Economics.

Gujarati, D.N. (2003). Basic Econometrics. New Delhi: Tata McGraw-Hill Publishing

Iarossi, G. (2006). The power of survey design. A user's guide for managing surveys. Interpreting results and influencing respondents. Washington, D.C.: The World Bank.

Kara, O. (2012). Researching for Hospitality and Tourism Management. An unpublished course material of the Uttarakhand Open University.

Kothari C.R. (2011). Research Methodology. UK: New Age International.

2.8 Possible Answers to SAEs

Answers to SAEs 1

1. Population can be classified into finite and infinite.

2. The smaller the sample error, the greater the precision of the estimate and the more acceptable the research result.

3. The study of the entire population could be possible in certain research studies under the following conditions:

- When the population of study is very small;

- Time available for the research is small;
- Material and human resources are inadequate;
- When complete population count is the primary objective of the study **(Any 3).**

4. The use of the population sample for a study is necessary when:

- The population size is very large;
- Time, human and material resources available are adequate.

Answers to SAEs 2

1. Sampling is the selection of the part of the population on the assumption that the part is the representative of the whole.

2. The individual and the type of research undertaken.

3. Sampling frame is a list of every member (or unit) of the population from which the sample will be drawn.

4. There are two basic requirements in sampling: a sample must be representative and it must be adequate (of sufficient size).

Unit 3: SAMPLING TECHNIQUES

Unit Structure

3.1 Introduction

A research may involve a very small population of study and in this case, all the members of the population can be included in the study with ease. Another research may focus on large population and so it is neither feasible nor economic to obtain information from all the members of the population. In this case, enumerating a population (that is, census) involves a great deal of time, fund and other resources, usually beyond the reach of researchers. Thus, it becomes imperative to collect information from only a part of the population. Sampling is therefore a process used in statistical analysis in which a predetermined number of observations are taken from a larger population. The methodology used to sample from a larger population depends on the type of

analysis being performed but may either be probability or non-probability sampling techniques as discussed in this unit.

3.2 Learning Outcomes

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

- Define sampling method or technique;
- State the types of sampling methods with relevant examples; and
- Analyse the different categories of sampling methods/techniques.

3.3 Sampling Techniques

3.3.1 Definition of Sampling Methods/Techniques

Sampling method or sampling design involves the procedure adopted in selection of the sample size to represent the target population of the research. It is a plan which specifies how some sample elements will be selected from a given population. This depends on the research objectives and purpose, the accessibility of the population and availability of sampling frame. The quality of a sample is assessed based by the procedure that produced it; and the ability to estimate and reduce sampling error depends on the sampling method used.

3.3.2 Types of Sampling Methods/Techniques

There are two broad categories of sampling methods. These are probability and non-probability sampling. Nonetheless, every researcher must take into consideration some important factors such

as availability of funds, population of study and accessibility of the elements or respondents while making his choice of the sampling technique to adopt for a particular study.

3.3.2.1 Probability sampling technique

Probability sampling methods involve random selection procedures to ensure that each unit of the sample is chosen on the basis of chance. All units of the population should have an equal or at least a known chance of being included in the sample. Invariably therefore, a probability sampling procedure requires a sample frame to be compiled or adapted. Though creating or obtaining a sampling frame is difficult, particularly in developing countries where data is generally weak and incomplete, a good sampling frame is the raw material for good probability sampling. In statistical analysis, probability sampling techniques are preferred to non-probability sampling because only probability or random sampling provides researchers with reliable estimates of the unknown population value under investigation.

The basic advantage of probability sampling methods is that they assure that the sample chosen is representative of the population thereby ensuring that the statistical conclusions stand valid. It minimizes the risk of sampling bias and it enables a researcher to draw inferences, from the sample' about the population from which the sample is drawn, with levels of confidence that can be estimated statistically. If the researcher knows the chances of including each population element in the sample, then he or she can use statistical theory to estimate the properties of the statistics obtained through the survey (Iarossi, 2006). Probability sampling methods or techniques include the following types of sampling:

- Simple random sampling;
- Systematic sampling;

- Stratified sampling;
- Cluster sampling; and
- Multistage sampling.

a. Simple Random Sampling

Simple random sampling is the simplest type of probability sampling. This refers to a sampling method where each member or unit in the population has the same probability of being selected in the sample. Here, selection is often made with the aid of random numbers after the units in the frame have been labelled from 1 to N and all the samples have the same probability and are therefore equivalent. For instance, given a list of all the English alphabets, the probability of drawing an alphabet is 1/26 while that of drawing a vowel from the list is 5/26. Thus, any simple random sampling technique should possess certain features such as:

- its population consists of N objects.
- its sample consists of n objects.
- if all possible samples of n objects are equally likely to occur, the sampling method is called simple random sampling.

This is a sampling procedure which ensures that every unit in the study population has a known and equal chance (that is, equi-probability) of being included in the sample. Since once chosen, a unit cannot be chosen again for the sample and given that choosing one unit for the sample does not preclude the chance of any other unit being chosen, a simple random sample of sufficient size is unlikely to over-represent or under-represent certain type of unit. With simple random sampling, every possible sample of equal size as well as every individual, element in the population has the same non-zero probability of being selected. The steps involved in drawing a simple random sample are:

- decide what is to be the unit of sampling; is it a place, household, person, couple?;
- make a list of all the units in the population or study universe from which the sample is to be drawn (that is, compile a sampling frame). This step is based on concrete description of the universe to avoid ambiguities and confusions. Where a sampling frame exists already, ensure that they are complete and up-to-date; and
- select the required number of sampling units (sample size) using a lottery method or a table of random numbers or by computer-generated random numbers.

Tables of random numbers are a simple and convenient source of randomized selection of sample units. The use of a table of random numbers, like the selection from a box, involves numbering serially every unit in the population so that every member has the same number of digits on it. Then numbers are read from a table of random numbers in any way desired up, down, horizontally or diagonally, and the corresponding units are selected for the sample. The sample size in simple random sampling (equal probability of selection design) depends on three factors: the size of the target population, the variability of the population parameter which the researcher wishes to estimate and the degree of precision and confidence aimed by the researcher.

b. Systematic Random Sampling

Systematic random sampling is a systematic selection of elements from a given list at fixed interval in a randomization manner. For example, in a list of 1000 students, a systematic random sample could be obtained by selecting every tenth person on the list. This is only possible when the population of study is finite and manageable. This method of sampling is a variation of the simple random sampling. Systematic random sampling is characterized by a random start, followed by a pre-determined or systematized order of selection. Once the first unit has been selected (that is, the random start), all the rest of the units for the sample are pre-determined. It is probability sampling

because the chance of selection for each unit is known and the starting point is random. This begins with creation of a list of each member of the population. From the list, we randomly select the first sample element from the first *k* elements on the population list. Thereafter, we select every *kth* element on the list. The procedures for a systematic random sampling are:

- a population to be sampled must be listed in some manner. It is not required that the units be individually numbered but the count of units must be known.
- total number of units in the sampling frame (N) is divided by the units in the sample (n) to obtain the sampling interval (K) rounded up to the nearest whole number. K = N/n. For example, if there are 500 students in a particular Department and a sample of 50 is to be drawn, then the sampling interval K is 500/50 = 10.

Systematic random sampling enables the researcher to obtain a sample that is evenly distributed in the population unlike simple random sampling. It saves time and costs of doing a list of large populations and still retains the features of probability sampling. Even though, for statistical reasons, simple random sampling is more likely to produce a study population that is representative, systematic sampling is in practice easier and quicker to do. The major disadvantage of systematic random sampling is that it is restricted only to homogeneous population. Another problem with systematic random sampling is when the sample frame is not in random order, that is, if the population list exhibits an intrinsic regulatory or periodicity of its own. This is otherwise referred to as systematic variation. For example, if a researcher is conducting a systematic sampling at sampling interval of 10 from the population list ordered from rich to poor, The selection process will be biased to include all the possible samples It would be impossible for such a procedure to include some households of all the different levels of size and as a result, systematic sampling can lead to bias.

Therefore, in proposing to use systematic random procedure to select a sample, a researcher must check whether the sampling frame has some in-built regularity of some sort, and if it has, it may be advisable to use simple random sampling or otherwise re-arrange the frame to do away with the periodicity. Where the frame is ordered or ranked overall in some way, for example, a list of the households in a community ordered by size so that the village with the largest population was the first on the list, followed by next largest, and so on, finishing with the smallest household at the end of the list, then the sampling frame would be satisfactory for use in systematic sampling. In fact, having a frame ordered in some way can increase the sampling efficiency. When dealing with heterogeneous population, it is better to first partition the entire population into homogeneous groups before the application of systematic random sampling technique.

Self-Assessment Exercises 1

1. There are two broad categories of sampling methods--- and ---?

2. What is systematic random sampling?

3. A Hospitality and Tourism Reservation Manager is conducting a satisfaction survey, sampling from a list of 5,000 new guests in various hotels. The list includes 1,000 guests from the Ebonyi Hotels, 1,000 guests from the Hope High Hotels, 1,000 guests from the Jejeson Hotels, 1,000 guests from the Jagua Techs Hotels, and 1,000 guests from the Steve World Glass Hotels. He selects a sample of 500 room guests, by randomly sampling 100 guests of each brand.

   Is this an example of a simple random sample? Choose from these options with reason (s):

a) Yes, because each guest in the sample was randomly sampled.

b) Yes, because each guest in the sample had an equal chance of being sampled.

c) Yes, because guest of each brand was equally represented in the sample.

d) No, because every possible 500 guest sample did not have an equal chance of being chosen.

e) No, because the population consisted of purchasers of five different brands of hotels.

c. Stratified Random Sampling

In stratified sampling technique, the units in the population are allocated to sub-groups or strata on the basis of resemblance or similarity. This make it possible to produce an estimate for the population characteristic of which is considerably better than that given by a simple random sample from the whole population under study. With stratified sampling, the population is divided into groups, based on some characteristic. Then, within each group, a probability sample (often a simple random sample) is selected. In stratified sampling, the groups are called **strata**. In this technique, the population is composed into strata (high, low, above, below) and then drawing a sample from each level; and ensuring that each stratum of the population is represented in the sample before randomization to select its members either by tossing of the coin or use of the table of random numbers of systematic sampling.

The purpose of stratified random sampling is to produce subpopulations whereby the units of each subpopulation are alike in one major characteristic. Stratification means dividing the study population into mutually exclusive sub-population called strata and drawing a sample in a random manner. The basic rationale for stratified random sampling is the fact that a homogenous population requires a smaller sample than does a heterogeneous sample. While simple random might lead to representation, stratification helps to achieve better representativeness of the population being studied. For instance, if we are to carry out a survey of all Universities in Nigeria we will likely begin with classification according to type of University: Public (Federal or state?), Private (Mission or individual ownership?), according to progammes, according to location and so on.

Many different types of stratification of variables or criteria may be used to divide the population into strata. For example, division of residents into different categories of urban and rural areas or different age, gender, income groups and so on. Strata created must be more homogenous than the population as a whole. By this reason, stratification reduces sampling variability and increases efficiency of sampling. If the stratification does not produce relatively homogenous groups, there is no advantage in its use. But if it does, then the precision of the overall estimates for a given sample size will improve substantially. Were a simple random sampling to be used instead, the sample size would have to be larger in order to obtain the same level of precision. According to Eboh (2009), the actual mechanics for drawing a stratified random sample is given below.

- Divide the population with appropriate sub-groups (strata), not necessarily of equal size;
- Considering each sub-group as a separate population draw a sample using some random type procedure such as proportional sampling, equal size sampling or optimum allocation sampling.

Proportional sampling (or proportional probability sampling, PPS) means that an equal percentage of each stratum will be selected, regardless of stratum size. Equal-size sampling means that an equal number of units are randomly selected from each stratum. For example, 50 units may be selected from each of three strata, regardless of their size, to yield a sample size of 150. But there has to be varying proportions in the separate strata by use of a weighting factor. If no weighting is done, then data should be presented separately for each group and not combined or pooled.

By this weighting, a researcher can use stratified sampling to take a relatively large sample from a small group in the study population, thereby allowing the researcher to get a sample that is big enough to enable valid conclusions to be drawn about a relatively small group without having to collect an unnecessarily large (and of course, expensive) sample for the other larger groups.

Optimum allocation sampling involves using formula to determine the sample size for each stratum that will yield the most precise estimates; also it is necessary to adjust for varying proportions of the sample in the separate strata, using a weighting factor.

d. Cluster Sampling

Cluster sampling is akin to agricultural economics and extension statistics. Here, the researcher divides the area into cluster of zones, creates units from each zone, randomly selects the units and individual elements within the units are sampled for study. This method of sampling divides the population into contiguous groups called clusters. Samples are then drawn from the clusters, by random method. Clusters are often geographic areas/units (for example, villages, districts) or organizational units (for example, farmers' cooperative society, projects clinics). Clusters can also be group of households or animals in a herd.

Suppose Ohaukwu Agricultural Area is to be surveyed, we first divide into zones on clusters of communities (Umuogudu-Oshia, Umuogudu-Akpu, Amoffia, Ekwashi, Ezzangbo, Effium, Okposhi-Eheku, Okposhi-Eshi, Umuezeaka and Ukwagba). Then each of these communities has a number of villages (units). For instance, Umuogudu-Oshia community is composed of villages such as Ekwebru, Ndiagugba, Ndiagudu, Okrogbata, Ogene, Abarigwe, Ndiaguobu, Adabru etc. Each of the aforementioned village is made up of hundreds and thousands of households (individuals). Simple random sampling technique is then applied to select a sample of individuals within the villages for the study.

Cluster Sampling is the selection of groups of study units (clusters) instead of the selection of study units individually. Where it is difficult or impossible to obtain a sample because of lack of a complete sampling frame and easily compiled and a random selection of these had done. Here, a sampling frame of the individual units is not required for the population. Only a sampling frame

of clusters is needed initially and then a sampling frame within individual clusters. Cluster sampling is suitable for use when a relatively common condition is being investigated or the focus of interest is the occurrence of individual events within a particular carefully specified locality; like in surveys investigating births, deaths or migration. The following steps are involved in drawing a sample by cluster method:

- construct a sampling frame of the clusters, villages, industries and so on, as may be required depending on the study.
- randomly select clusters to be included in the study; the number of clusters depends upon the needs of the research. Studies which measure numerous variables, some of which may be less common should include larger sample sizes.
- for each of the clusters included, draw the units of sampling in a random manner. This may mean constructing a sampling frame for each of the selected clusters and using a simple random or systematic random method to draw the sample from each cluster.

Despite the logistic and practical advantage of cluster sampling, there is usually a tendency for units in the same cluster sampling, to be comparatively more similar to each other than to units in different clusters, especially where clusters are specific localities. This tendency is measured by intra-cluster correlation. If the intra-cluster correlation coefficient for a variable is positive, then the sampling clusters or groups of units are less efficient than simple random sampling of units spread evenly over the whole population. Consequently, the likelihood of the sample not being representative of the whole population depends largely on the number of clusters selected in the first page.

The larger the number of clusters, the greater the likelihood that the sample will be representative. The larger the size' of the cluster the less efficient the sample is. So, while stratification will almost invariably produce a more precise estimate than a simple random sample, cluster sampling

generally gives a less precise estimate. Cluster sampling is still considered a probability sample because the probability of selecting a unit at each stage of the sampling (the sampling fraction) can be determined and thus the sample precision estimated. With cluster sampling, every member of the population is assigned to one, and only one, group. Each group is called a cluster. A sample of clusters is chosen, using a probability method (often simple random sampling). Only individuals within sampled clusters are surveyed. With stratified sampling, the sample includes elements from each stratum. With cluster sampling, in contrast, the sample includes elements only from sampled clusters.

e. Multi-Stage Sampling

Multi-stage sampling is a combination of two or more other sampling methods. In this case the process of estimation can be carried out stage by stage, using the appropriate methods of estimation at each stage. This method of sampling involves a procedure whereby the selection of units into the sample is organized into stages. It usually involves a combination of sampling methods. In this method of sampling, we select a sample by using combinations of more than one sampling method. For example, in Stage 1, we might use cluster sampling to choose clusters from a population. Then, in Stage 2, we might use simple random sampling to select a subset of elements from each chosen cluster for the final sample.

Multi-stage sampling can be described as follows: a first stage in which a sample of larger units is selected. Then a second stage in which, from each of the selected first stage units, a sample of smaller units is selected. Further stages may be involved as appropriate. The instance given above is a two-stage sampling method. By the method, a researcher draws a sample of, say, villages or

other settlement type or locality from a list of villages the first stage. Then, the researcher proceeds to a second stage and draws a sample of households within each village.

There can be a three stage sampling whereby a researcher draws a sample of districts within a region (first stage), villages within each selected district (second stage) and lastly a sample of households within each selected village (third stage). In four stage sampling, a researcher can select local government areas from each senatorial zone, village and farm plots from each farmer in the sample. The shared attribute of multistage sampling is that it is possible to reach final sampling or study units through a hierarchy of higher level stages. The variance of the estimate in multi-stage sampling is built up of two components: a variation between primary sampling units and the variation within primary sampling units. Sum of these two components gives the variance of the estimate. The advantage of multi-stage sampling is that frame has to be prepared for only those units from which samples are to be selected thereby saving survey cost.

However, some difficulties arise in multi-stage sampling. Suppose a situation where the primary sampling unit is villages and the secondary sampling unit is households within villages, then for the design to be self-weighting it is necessary that the overall probability of selection of each household should be the same, no matter which village the household belongs to. Two main approaches can be used to handle this problem. One is sampling with constant sampling fraction. The other is sampling with probability proportional to size.

3.3.2.2 Non-probability or non-random sampling technique

Non-probability sampling technique does not operate within the process of randomization principle because the researcher selects his sample with bias, interest or favoritism. The selection method here is not based on probability theory; and as such a true representation of the population

may not be obtained and the degree to which the sample differs from the population may remain unknown. Where a sampling frame is not available and cannot be improvised perhaps for logistic reasons, it may not be possible to sample the study units such that the probability for the different units to be selected is known. In such situations, non- probability sampling methods are used; the sample is selected not by chance but for reasons of practical convenience, personal judgment, and subjective inclusion or in order to include units with particular characteristics in the sample. This procedure should be used with caution and perhaps only where probability sampling is not practicable, otherwise the representativeness of the sample may be low and statistics computed could be of poor reliability (Gujarati, 2003). While all non-probability samples are not equally subject to bias, some might be more biased than others. The main drawback of non-probability sampling is that, in contrast to probability sampling, the researcher cannot apply statistical theory to examine the properties of sample estimates because non-probability samples can be assessed only through subjective valuation.

Generally, if the aim of research is to measure variables and generalize findings obtained from a sample to the whole study population, then non-probability sampling methods are inappropriate. Nevertheless, non -probability sampling is most likely to be used in rapid or exploratory studies where information is needed very quickly but not with great precision and also in case studies. This does not necessarily mean that the use of non-probability sampling is limited to such studies.

Types of non-probability sampling methods include:

- Convenience (judgment or opportunity) sampling ;
- Purposive sampling; and
- Quota sampling.

a. Convenience Sampling

As the name suggests, this is a sample selection based on convenience. A convenience sample consists of people who are easily approachable and can be reached out to in shorter time. This is a method of non-probability sampling in which a researcher selects for the sample size those study units that happen to be available at the time of data collection which he happens to come across. The overriding principles and criteria here are researcher convenience and respondent availability. The risk of bias in such a method is obvious; for example, persons encountered in a particular village by a researcher may not represent the characteristics of the whole people in the village.

In convenience sampling, a major weakness is that some units may be over-selected, others under-selected or missed outright that is, some characteristics of people in the village may be over-represented, others under-represented or missed outright. The convenience sampling method is often used during preliminary research efforts to get a gross estimate of the results without incurring the maximum cost of resources and time required to obtain an appropriate random sample. Here the researcher is just interested in getting an inexpensive approximation of the truth about the population of study.

b. Quota Sampling

Quota sampling involves extracting a representative portion of elements, objects or individuals with a particular characteristic from a population; by sub-grouping the population according to identifiable characteristics and hand-picking some predetermined number from each sub-group. This approach enables the researcher to admit any category of traits, qualities or elements that are of particular interest to him into the sample. This method of sampling could yield more representative sample than convenience sampling. This is because even though in general a

researcher has considerable freedom of choice, there is the restriction that certain characteristics (for example, age, sex, socioeconomic status, and so on) of the respondents must be typical or representative of the area of the group being studied. By this conditionality, quota sampling ensures that a certain number of sample units from different categories with specific characteristics appear in the sample so that all these characteristics are represented.

Quota sampling procedure is often helpful in minimizing bias in non- probability sampling. This is achieved by establishing quotas for different groups in the population and then followed by convenience or purposive sampling to fill each quota. So, a researcher interviews as many people in each category as he can find until the given quota is filled. In spite of the advantages of quota sampling over other non-probability methods, quota sampling is not a satisfactory procedure if high precision is required since there is no way of calculating the limit of permissible error. Even though, quota sampling has been utilized because of its relative inexpensiveness, and for the fact that it often yields rough results that may be satisfactory for the purpose desired, it cannot be said to yield a sample that is representative of the entire population. Although this technique is cheap, fast and easy to apply, it yields a highly biased sample because most of the population characteristics me be entirely absent in the composition. Only when quick and crude results will satisfy, is quota sampling successful, in the practical sense as the sample so composed cannot be said to be a true representative of the population of study.

c. Purposive Sampling

This is sampling with a particular purpose in mind. It draws a sample to illustrate or represent some particular characteristic in the population that meet certain criteria. This can be very useful for situations where the researcher needs to reach a targeted sample quickly and timely. For

instance, selecting only respondents who have adopted a certain innovation or technology in the sample. Purposive sampling is particularly likely to be found where a detailed study must cover a range of cases from different sub-populations such as farms in different ecological areas or respondents in different cultural areas.

A purposive sample can be based on the capacity of the respondent to supply the required information. A sample of areas or persons can be chosen purposively that is, because they have certain characteristics believed to be "typical". Just like any other non-probability sampling method, purposive sampling procedure is practically weak because the sample varies in unknown ways from the universe. A good example of purposive sampling is Snowball sampling. Snowball sampling relies on referrals from initial subjects to generate additional subjects. Here you begin with people who meets the criteria for inclusion in the study thereby reducing the research cost dramatically. Snowball sampling is especially useful when trying to reach populations that are either inaccessible or very difficult to come by. For illustration, in a study involving the childless couples within a specific geographical location; you have to search for couples in this category and still ask them about other childless couples in their vicinity and how to reach them.

Self-Assessment Exercises 2

1. In stratified sampling, the groups are called …?
2. What is multi-stage sampling?
3. State the three types of non-probability sampling methods.
4. Snowball sampling is especially useful when …?

3.4 Summary

The major objective of the researcher always should be to obtain a sample that is as much representative of the population as possible. However, this aim is not always easy to realize in practice due to several practical limitations in terms of time, funds and insufficient knowledge about the population under consideration. In light of the forgoing, every researcher has to take into account certain obstacles or factors while making decisions relation to research. In this units, we have been able to define sampling methods or techniques. Besides we described two major types of sampling: probability and non-probability sampling in details.

3.5 Glossary

Quota sampling: implies a predetermined number of what is desired

Panel: is a group of individuals who have agreed to provide information to a researcher over a period of time.

Random sampling: This is a sampling procedure which ensures that every unit in the study population has a known and equal chance of being selected and included as a sample".

3.6 References/Further Readings

Eboh, E.C. (2009). Social and economic research: Principles and methods (2nd Edition). Enugu: African Institute, for Applied Economics.

Gujarati, D.N. (2003). Basic Econometrics. New Delhi: Tata McGraw-Hill Publishing

Iarossi, G. (2006). The power of survey design. A user's guide for managing surveys. Interpreting results and influencing respondents. Washington, D.C.: The World Bank.

Kara, O. (2012). Researching for Hospitality and Tourism Management. An unpublished course material of the Uttarakhand Open University.

3.7 Possible Answers to SAEs

Answers to SAEs 1

1. These are probability and non-probability sampling.

2. Systematic random sampling is a systematic selection of elements from a given list at fixed interval in a randomization manner.

3. The correct answer is (D). A simple random sample requires that every sample of size *n* (in this problem, *n* is equal to 500) has an equal chance of being selected. In this problem, there was a 100 percent chance that the sample would include 100 guests of each brand of hotel. The fact that each guest in the sample was randomly sampled is a necessary condition for a simple random sample, but it is not sufficient. Similarly, the fact that each guest in the sample had an equal chance of being selected is characteristic of a simple random sample, but it is not sufficient. The sampling method in this problem used random sampling and gave each guest an equal chance of being selected; but the sampling method was actually stratified random sampling. The fact that hotel guests of each of the brands were equally represented in the sample is irrelevant to whether the sampling method was simple random sampling. Similarly, the fact that population consisted of guests of different hotel brands is irrelevant.

Answers to SAEs 2

1. In stratified sampling, the groups are called strata**.**
2. This method of sampling involves a procedure whereby the selection of units into the sample is organized into stages.
3. Types of non-probability sampling methods include:
    - Convenience (judgment or opportunity) sampling ;
    - Purposive sampling; and
    - Quota sampling.
4. Snowball sampling is especially useful when trying to reach populations that are either inaccessible or very difficult to come by.

MODULE 3 RESEARCH PROCESS

UNIT 1 Meaning and Stages of the Research Process

UNIT 2 Research Planning and Execution

Unit 1: Meaning and Stages of the Research Process

Unit Structure

1.1 Introduction

The research process involves at least seven steps or stages which must be followed sequentially in order to simplify the research procedure and eliminate confusion. These steps include: problem definition, objectives of the research, research design, data collection, data analysis, interpretation and validation of results. *This unit focuses on the discussion on the various stages for conducting a successful scientific research.*

1.2 Learning Outcomes

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

- Discuss the research process.
- Analyse the sequential steps or stages of the research process; and
- State the roles of research in hospitality and tourism management.

1.3 Meaning and Stages of the Research Process

Increasing emphasis upon research design is one of the salient trends that characterize hospitality, tourism management and related fields in recent years. Even scholars who are not directly involved in this disciplines are showing much interest in hospitality and tourism data obtained through field investigation methods as can be seen in this unit.

1.3.1 **Meaning of Research Process**

Every scientific activity follows a set of systematic and sequential steps. This set of sequential steps constitutes a research process. Research process involves the sequence of steps to be followed in executing a research project. According to Eboh (2009), the research process involves a

sequence of mutually-inclusive and re-enforcing steps and procedures for acquiring reliable knowledge. As a process, research is a sequence of activities or steps leading to an end product. It has a procedural and step-wise character. This provides a systematic procedure or plan of action to be adopted by researchers while carrying out a research such as: identifying a problem, defining the problem in precise terms in unambiguous terms, stating the problems in form of research questions and research hypotheses, designing the research, collecting data, analyzing data, interpretation of data, drawing conclusion and making recommendations based on research findings.

1.3.2 Stages of the Research Process

The major stages in the research process are:

1.3.2.1 Identification of the problem

1.3.2.2 Review of related literature

1.3.2.3 Statement of research hypotheses, research objectives and research questions.

1.3.2.4 Designing research

1.3.2.5 Data Collection

1.3.2.6 Data Analysis

1.3.2.7 Results Generalisations and Interpretation

1.3.2.8 Validation of Results.

1.3.2.1 Identification of the Problem

This is the first and most difficult step in the research process because all other steps will depend on the stated research problem. The research problem must be identified and defined accordingly without any ambiguity. The purpose of stating the research problem is to set out the scope of the study by precisely specifying the variables or areas to be investigated in the research project. This keeps the researcher focused and coordinated in his effort. A research problem could emanate from various sources such as research reports, published articles, textbooks, personal observations

and experiences. An identified problem is often described in a declarative statement. This implies stating a research topic in a precise, concrete and unambiguous manner. The research problem could relate to the state of nature or relationships between existing variables. Before a problem could be investigated, it must significantly present a state of knowledge and the research into the problem must be feasible with possibility of available relevant data. A good research problem should possess the following qualities:

- it must involve important issues as to necessitate an investigation;
- this issue can be solved through the scientific method;
- it must involve an issue that has not been conclusively researched on earlier;
- it must be executed within the possible time and available human and material resources.

1.3.2.2 Review of related literature

When the researcher identifies a research problem, he goes further to consult relevant literature connected to his area of study. Literature review is a critical examination of existing works in a field under research. This involves reading and studying academic materials related to the research problem which enables the researcher to learn about related problem in the past and how the problem was solved. Related literature has to do with having a look at the reports of what others have done in related areas of study so as to identify the similarities and differences existing between the findings of various researchers. The researcher brings out important facts of previous studies while eliminating findings of unrelated research problems. In the review of related literature, attempt is made to identify, locate, read and evaluate results of previous studies so as to observe ideas, opinions and inferences relevant to his proposed research. This affords the researcher the opportunity to gain in-depth knowledge of the area he intends to delve in, the

problems that have already been studied, the ones that are yet to be investigated and the techniques to be adopted for the investigation in view. This enables the researcher to familiarize himself with the major variables, theories, models, ideas, concepts, methods and techniques of the study. Chukwuemeka (2002) contends that literature review serves two purposes in the research report:

i. to set theoretical base for the research and;
ii. to set the current research into perspective to show the state of art.

Review of related literature guides the researcher towards selecting a new dimension appropriate for his study and also helps him to locate the methodology applied in previous studies that may be suitable for his study. It also provides a historical and chronological order of research initiatives completed and new areas that ought to be examined in the light of recent developments, expectations and demands. Literature review enables the researcher to build a body of new knowledge, new problems, and new directions in a particular area for his own use and for others interested in the same reference area covered (Ominyi & Odo, 2006). It launches the researcher to improved ways of acquiring knowledge and information from internet browsing, sifting, journals, articles, radio and television reports, books, magazines, conferences proceedings, periodicals, monographs, seminar paper, newspapers, government reports etc. Sources of literature review are basically classified into two: preliminary sources (such as index, abstract and catalogue) and main or primary sources like journals, conference papers, students' projects, final reports of original research investigations, theoretical analysis, periodicals, magazines, newspapers, students' projects, dissertations, Encyclopedias and Dictionaries.

It is noteworthy that the review of related literature should not be composed of bibliographical summaries of quotations, or be a recital of who says what paragraph by paragraph. Generally, a

short and artfully integrated review of the resources most relevant to the purpose at hand should be enough to convince other scholars that the researcher is sufficiently knowledgeable and capable of carrying out the research plan. The main essence of review of related literature is to identify areas that could guide the researcher to make contributions in his own research, when and where gaps are identified. This involves locating, reading and evaluating reports of research, casual observations and opinions that are related to the envisaged study through identification, selection and use of materials for the review. The importance of literature review include:

i. it provide information on present state of knowledge in the area of study;

ii. it provides information about possible and statistical procedures;

iii. it detects area of deficiency to avoid errors already committed;

iv. make comparison between past and current research possible;

v. helps to delimit the scope and size of the research study;

vi. aids the selection of more valid research problems;

vii.it provides skills in reading, interpreting and evaluating previous research findings relevant to the present study.

1.3.2.3 Statement of Research Hypotheses, Research Objectives and Research Questions

At this stage, the researcher finalizes the research questions, hypotheses and boundary of the study. The cardinal and specific objectives of the study must be taken into account. Research questions present the problems to be resolved in the study and must relate to the purpose, place, and present state of issues in the study area and the possible means of solving the research problem. Appropriate answers to the research questions provide a starting point to crystallize the objectives

and hypotheses of the research problem. Hypothesis is a tentative assumption made for the purpose of testing its logical validity or empirical manifestations. It is an assumption about a population of the study formulated for a situation where the inference is not explicit. Working hypothesis may be true or false but the correct fact can be ascertained only after collection and analysis of data relating to the study. Hypothesis enables the researcher to focus on the specific area of inquiry thereby keeping him on track. Besides research questions and hypotheses, the researcher should clearly define the boundary of the study. This considers the scope of the study in terms of size, complexities and constraints.

1.3.2.4 Designing research

This is the conceptual structure upon which the research would be executed. It represents the plan, structure and strategy for investigation, conceived to obtain answers to the research questions and to control variables in the study. The research design provides a complete guidelines for data collection. Generally, the essence of a research design include: selection of research approach, design of sampling plan, design of experiment and questionnaire. Based on the type of research under consideration, the researcher decides on the type of study to undertake which may be descriptive or experimental designs. Both the research approach and sampling design had been discussed in details in module 2. Each study involves different response variables. To design a suitable experiment for a given study, the researcher should first identify the response variables of the study. Then, identify the factors affecting each response variable, form the skeleton, write the model and define the components of the experiment.

A questionnaire is a data gathering instrument in which target respondents are given standardized or uniform questions. This contains a list of questions relating to the objectives of the study and the hypotheses to be verified. It consists of a set of well-formulated questions relating to the

objectives of the study which enables the researcher to probe and obtain responses from respondents. Questionnaires are of two types: structured (here the respondent is restricted to some response options) and the unstructured or open ended (where the respondent is free to supply responses in his own words). A good questionnaire possess the following characteristics:

i.it must be related to the study objectives

ii.it must yield consistent responses;

iii.all instructions must be clearly stated;

iv.it must be easy to understand, interpret and complete;

v.its responses must be quantifiable;

vii.its choice of language should be simple.

However, the questionnaire for the study must contain provisions to collect all information required for testing the stated hypotheses for the study. Thus, the generalized steps for designing questionnaires include:

- identify the research issues and finalise the set of hypotheses;
- formulate a set of questions for each issue;
- decide the question wordings based on its types;
- arrange the questions in appropriate sequence;
- pre-test the questions; and

- review the questionnaire for improvements.

On the pattern of questionnaire distribution, the researcher can either deliver the questionnaire personally to the respondent or then wait to collect the response immediately; he can repeat the visit to collect the responses; or send the questionnaire and obtain the responses through correspondence.

Self-Assessment Exercises 1

1. Define the research process.
2. Mention any five sources of research problem.
3. Literature review serves two purposes in the research report… and ….?
4. Outline four essence of research design.
5. How do you define a questionnaire?

1.3.2.5 Data Collection

Data is simply defined as the recorded observation made on the sample of study which may be discrete or variable. Data collection is indispensable in every research because it enables the researcher to answer the research questions and test the hypotheses he has formulated. Research data can be classified into primary and secondary data and its collection involves observation, experimentation and field trips. Data collection methods to be adopted on a given research depends on factors such as the type of research problem, the stated hypotheses, time and other resources at the researcher's disposal. The methods used for collecting primary data are personal observation method, personal interview, telephone interview, and mail survey. Secondary data on the other hand are collected from existing internal or external sources. The internal sources of secondary data collection include sales records, marketing activity, cost information, distributor reports,

feedback and customer feedback. External sources of secondary data collection are national government publications, foreign government reports, journals, trade association reports, books magazines, newspapers, annual reports, and scholarly research reports.

1.3.2.6 Data Analysis

Data collected during research do not become useful until they are analysed. This is the breaking down of recorded research information into meaningful components for critical examination, extraction of otherwise hidden information and making of interpretation and inferences about the relationships existing among the parts. Data analysis involves ordering, categorizing, manipulating and summarizing data, the aim of which is to reduce it in such a way as to render them useful in answering research questions and testing stated hypotheses. Data analysis refers to those techniques whereby the researcher extracts from the data information that where not apparently there before and which would enable a summary description of the subject studied to be made. Providing answers to research questions and working hypotheses would remain impossible unless the collected data for the study have been scrutinized using systematic and verifiable statistical techniques. This requires several closely related operations including data organization, exploration, tabulation and statistical testing of working hypotheses for the study. Descriptive and inferential statistics are the tools used for data analysis. Statistics on its own is defined as the science of collection, presentation, analysis and interpretation of numerical data.

Descriptive statistics is a branch of statistics which describes characteristics of a group under investigation. It describes a typical score in a group of scores or the extent of differences among the scores in a group. They are used for descriptive studies through the analysis of questionnaire data. Examples of descriptive statistics include percentage and frequency distribution, measures

of central tendency/location (mean, mode, median); measures of variability/dispersion/ spread (range, mean deviation, variance, standard deviation); measures of relationships and measures of relative understanding. These descriptive statistics enables the researcher to describe observed events, explain their order of importance, state the extent two sets of values of scores are related and the relativity of individual performances

Inferential statistics which is also called the parametric data analysis is a statistical tool for establishing the degree of association or relationship between two or more variables. It provides a means for evaluating the relationship that exists within the data collected from a sample of the population. Examples of inferential statistical tools are Correlation coefficient, Regression analysis, Chi-square, T-test, Z-test, Analysis of Variance (ANOVA), Analysis of Covariance (ANCOVA). Inferential statistics enables the researcher to test formulated hypotheses and make generalization or inferences to the population a particular sample was drawn from. Inferential statistics are more rigorous than descriptive statistics in their application to data analysis. Generally, data analysis involves the following tasks:

- computation of statistics (descriptive and inferential),
- designing regression equation for estimating response variable as a function of set of independent variables,
- performing correlation analysis,
- testing different hypotheses relating to various issues of research,
- factor analysis,
- discriminant analysis,
- conjoint analysis

1.3.2.7 Results Generalisations and Interpretation

The basic significance of research is that it seeks to arrive at generalisations based on empirical observation and logical validity .The analysis of data after data collection yields a set of results either in the form of statistics, regression equations, identification of significant factors or in the form of acceptance or rejection of different hypotheses. The researcher must infer the results of the original research issues from the results obtained through data analysis by taking various assumptions of the study into consideration. This involves explaining the results of research in the context of existing theories or previous studies. Through generalization and interpretation, new research questions could be established that may even require further investigative research.

1.3.2.8 Validation of Results.

A test is said to be valid if it purports to do what it is supposed to do. The essence of validation is to provide some evidence on the basis of which such inferences can be substituted. Validation aims to find whether the results satisfy the criteria of magnitude and logic with respect to the correctness of the data collected. The results of a given research study after interpretation must be validated by using past data. The process of validation of results ensures the credibility of the results. If there is any mismatch between the results of the model and that of the real-world problem in the past, then the assumptions and modelling exercise should be revisited till the results are validated. In the case of testing hypotheses, the concept f validation has different meanings. The credibility of the inferences of testing of hypotheses are already built-in through a significance level. Inferences from the research can act as guidelines for framing policies or some correction/change in the system of study for further betterment (Panneerselvam, 2013).

1.4 Roles of Research in Hospitality and Tourism Management

i. Research enables the creation of new knowledge thereby extending the frontiers of knowledge in hospitality and tourism management as a discipline;

ii. Research provides sustainable solution to the problems of hospitality and tourism in Nigeria. It brings about improved administrative achievements in the discipline.

iii. Through research better and innovative tools and techniques are developed for the improvement of service delivery to the people;

iv. Research also enables proprietors to take advantage of the various human, administrative and social theories to improve their management skills and practice.

v. Research leads to continuity and general growth, development and improvement in the field of hospitality and tourism management.

vi. It ensures continuous future growth of the hospitality industry and makes the proprietors professionally competent.

vii. Through research breakthroughs, unprofitable methods are discarded while more innovative approaches to the hospitality business are adopted for greater efficiency and profitability.

Self-Assessment Exercises 2

1. Mention four methods used for collecting primary data.
2. Define statistics.
3. Give any four examples of inferential statistics.
4. A test is said to be valid if …?
5. List the three examples of measures of central tendency.

1.5 Summary

The data collected from the samples is arranged in meaningful way by editing, coding, and presented in tabular form for drawing useful inferences. In this unit we have learnt the various ways by which raw data is converted into important information through the stages of research process such as identifying a problem, defining the problem in precise terms in unambiguous terms, stating the problems in form of research questions and research hypotheses, designing the research, collecting data, analyzing data, interpretation and validation of research results as discussed in details here.

1.6 References/Further Readings

Eboh, E.C. (2009). Social and economic research: Principles and methods (2nd Edition). Enugu: African Institute, for Applied Economics.

Gujarati, D.N. (2003). Basic Econometrics. New Delhi: Tata McGraw-Hill Publishing

Panneerselvam, R. (2013). Research methodology. Dehli: PHI Learning Private Limited.

1.7 Possible Answers to SAEs

Answers to SAEs 1

1. Research process involves the sequence of steps to be followed in executing a research project.

2. A research problem could emanate from various sources such as research reports, published articles, textbooks, personal observations and experiences **(Any 5).**

3. Literature review serves two purposes in the research report:

i.to set theoretical base for the research and;

ii. to set the current research into perspective to show the state of art.

4. The essence of a research design include: selection of research approach, design of sampling plan, design of experiment and questionnaire.

5. A questionnaire is a data gathering instrument in which target respondents are given standardized or uniform questions.

Answers to SAEs 2

1. The methods used for collecting primary data are personal observation method, personal interview, telephone interview, and mail survey.

2. Statistics is defined as the science of collection, presentation, analysis and interpretation of numerical data.

3. Examples of inferential statistical tools are Correlation coefficient, Regression analysis, Chi-square, T-test, Z-test, Analysis of Variance (ANOVA), Analysis of Covariance (ANCOVA).

4. A test is said to be valid if it purports to do what it is supposed to do**.**

5. Measures of central tendency include: mean, mode and median.

Unit 2: Research Planning and Execution

Unit Structure

2.1 Introduction

Planning and execution are two important parts of the problem-solving process in research. Based on related research, it is expected that planning speed and execution speed are positively correlated because of underlying individual differences in general mental speed. While there could also be a direct negative dependency of execution time on planning time, given the hypothesis that an investment in planning contributes to more efficient execution as can be seen in this unit.

2.2 Learning Outcomes

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

- Explain the meaning of research planning;
- Write short notes on research execution, and
- Analyse the steps research planning and execution.

2.3 Research Planning and Execution

1.3.1 Meaning of research planning

Planning research projects is a time-honoured intellectual exercise: one that requires both creativity and sharp analytical skills. Research planning involves broadly similar steps, including: formulating a question, developing an argument or predictions based on previous research, then selecting the information needed to answer your question. At its simplest, research planning involves the four distinct steps outlined below: orienting yourself to knowledge-creation; defining your research question; reviewing previous research on your question; and then choosing relevant data to formulate your own answers. Because the focus of this unit is on planning a research project, as opposed to conducting a research project, this section won't delve into the details of data-collection or analysis; those steps happen after you have planned the project.

1.3.2 Meaning of research execution

The execution phase is the period of active data collection and analysis. If your field work is being supported by human resources, it is recommended that you establish an employment contract in writing prior to engaging staff in thesis support work. In the contract specify:

- The nature of the work to be done
- The work hours
- Deliverables
- Supervision and quality assurance procedures (if necessary)
- Compensation
- Travel Allowances (if necessary)

Both the researcher and the contracted resource person should sign the document and each should keep a copy. In the event of future disputes about work requirements or compensation, each can refer to the signed contract. A written agreement should be established for all human resources supporting your thesis research, including: field assistants, laboratory assistants, translators, transcribers, and drivers. It is recommended that all field assistants undergo a training period prior to beginning to collect data. Depending on the complexity of the study, a training manual may be required. Consult your supervisor if in doubt about the need for a training manual. If you are acquiring lodging in the field during your data collection, it is recommended that you also secure a contract stating the agreed upon fee (daily, weekly or monthly as the case may be) and signed by both parties before taking up residence.

After hiring and training field assistants, one of the first activities you should undertake is to field test your data collection instruments – this applies to both quantitative (e.g. surveys) and qualitative instruments (e.g. interview guides). If your field location is in foreign country, it may be required to translate your data collection instruments into the local language. Ask your local

partner institution for advice on this. When translating from one language to another, it is recommended to have one person translate it from the original language to the local language and another person to translate it back to the original. This will allow you to determine if the translation is of a sufficient quality. As a researcher, you should test your survey or guides on one or two participants to determine if the questions, phrasing, order and tone are appropriate. Based on your experience with the instrument testing, you will make adjustments as needed. Quantitative instruments should not change after this initial field test. Qualitative instruments may continue to evolve throughout the interview period based on your reflection and field observations.

Participant Recruitment

During the course of developing your thesis proposal, you will have defined an approach to participant recruitment. For quantitative field studies, this will have included: participant inclusion criteria, exclusion criteria and sample size. For qualitative studies, your approach to participant recruitment should also have been defined. There are a number of qualitative approaches that may be utilized, including: purposive, convenience and snowball sampling. For desk reviews, you should also define the parameters of your literature search. During the execution phase, these parameters may be refined to expand or limit publication dates, geographies, or study types based on your initial search results. When recruiting and collecting data from participants, be sure to present them either in writing or verbally with an approved consent form. If necessary, your consent form should be translated into the local language.

Be sure to check the requirements of your home university, partner institutions and local authorities around ethical clearance applications. In many cases, data collection cannot begin until ethical clearance has been secured and ethical clearance is usually a requirement for publishing findings in peer reviewed journals. The ethical clearance process can take up to 3 months to complete and sometimes requires multiple efforts. Be sure to plan accordingly. Please note that even if formal ethical clearance isn't required, if you are collecting data in a low or middle income country it may be customary and necessary to meet with local village chiefs, local assemblymen or other important stakeholders to inform them of your research work and seek their approval to collect data (Brooke, 2022)..

Self-Assessment Exercises 1

1. The execution phase of research is the period of …?
2. State the steps involved in research planning.
3. Ethical clearance is usually a requirement for …?
4. Mention any five issues to be considered while engaging staff in thesis support work.

2.4 **Steps in research planning and execution**

Zhaojun, Paul, and Jian (2020) identified four steps in research planning and execution for researchers, namely: orient yourself, define your research question, review previous research and choose your data and methods.

i. Orient yourself

Planning and conducting research requires you to make a transition, from thinking like a consumer of information to thinking like a producer of information. That sounds simple, but it's actually a complex task. As a practical matter, this means putting aside the mindset of a scholar, which treats knowledge as something created by other people. As scholars, we are often passive receivers of knowledge: asked to do a specified set of readings, then graded on how well we reproduce what we've read. Researchers, however, must take on an active role as knowledge producers. Doing research requires more of you than reading and absorbing what other people have written: you have to engage in a dialogue with it. That includes arguing with previous knowledge and perhaps trying to show that ideas we have accepted as given are actually wrong or incomplete. For example, rather than simply taking in the claims of an author you read, you'll need to draw out the implications of those claims: if what the author is saying is true, what else does that suggest must be true? What predictions could you make based on the author's claims?

ii. Define your research question

Students often give this step cursory attention, but experienced researchers know that formulating a good question is sometimes the most difficult part of the research planning process. That is because the precise language of the question frames the rest of the project. It's therefore important to pose the question carefully, in a way that's both possible to answer and likely to yield interesting results. Of course, you must choose a question that interests you, but that's only the beginning of what's likely to be an iterative process: most researchers come back to this step repeatedly, modifying their questions in light of previous research, resource limitations and other considerations.

Researchers face limits in terms of time and money. They, like everyone else, have to pose research questions that they can plausibly answer given the constraints they face. For example, it would be inadvisable to frame a project around the question 'What are the roots of the Arab-Israeli conflict?' if you have only a week to develop an answer and no background on that topic. That's not to limit your imagination: you can come up with any question you'd like. But it typically does require some creativity to frame a question that you can answer well – that is, by investigating thoroughly and providing new insights – within the limits you face.

In addition to being interesting to you, and feasible within your resource constraints, the third and most important characteristic of a 'good' research topic is whether it allows you to create new knowledge. It might turn out that your question has already been asked and answered to your satisfaction: if so, you'll find out in the next step of this process. On the other hand, you might come up with a research question that hasn't been addressed previously. Before you get too excited about breaking uncharted ground, consider this: a lot of potentially researchable questions haven't been studied for good reason; they might have answers that are trivial or of very limited interest. This could include a question such as 'Why does the area of a circle equal πr^2? Finally, scholarly research questions must in some way lead to new and distinctive insights. For example, lots of people have studied gender roles in sports teams; what can you ask that hasn't been asked before? Reinventing the wheel is the number-one no-no in this endeavour. That's why the next step is so important: reviewing previous research on your topic. Depending on what you find in that step, you might need to revise your research question; iterating between your question and the existing literature is a normal process.

iii. Review previous research

In academic research, from articles to books, it's common to find a section called a 'literature review'. The purpose of that section is to describe the state of the art in knowledge on the research question that a project has posed. It demonstrates that researchers have thoroughly and systematically reviewed the relevant findings of previous studies on their topic, and that they have something novel to contribute. Your own research project should include something like this, even if it's a high-school term paper. In the research planning process, you'll want to list at least half a dozen bullet points stating the major findings on your topic by other people. In relation to those findings, you should be able to specify where your project could provide new and necessary insights. There are two basic rhetorical positions one can take in framing the novelty-plus-importance argument required of academic research:

- Position 1 requires you to build on or extend a set of existing ideas; that means saying something like: 'Person A has argued that X is true about gender; this implies Y, which has not yet been tested. My project will test Y, and if I find evidence to support it, that will change the way we understand gender.'
- Position 2 is to argue that there is a gap in existing knowledge, either because previous research has reached conflicting conclusions or has failed to consider something important. For example, one could say that research on middle schoolers and gender has been limited by being conducted primarily in coeducational environments, and that findings might differ dramatically if research were conducted in more schools where the student body was all-male or all-female.

Your overall goal in this step of the process is to show that your research will be part of a larger conversation: that is, how your project flows from what's already known, and how it advances,

extends or challenges that existing body of knowledge. That will be the contribution of your project, and it constitutes the motivation for your research.

Two things are worth mentioning about your search for sources of relevant previous research. First, you needn't look only at studies on your precise topic. For example, if you want to study gender-identity formation in schools, you shouldn't restrict yourself to studies of schools; the empirical setting (schools) is secondary to the larger social process that interests you (how people form gender identity). That process occurs in many different settings, so cast a wide net. Second, be sure to use legitimate sources – meaning publications that have been through some sort of vetting process, whether that involves peer review (as with academic journal articles you might find via Google Scholar) or editorial review (as you'd find in well-known mass media publications, such as The Economist or The Washington Post). What you'll want to avoid is using unvetted sources such as personal blogs or Wikipedia. Why? Because anybody can write anything in those forums, and there is no way to know – unless you're already an expert – if the claims you find there are accurate. Often, they're not (De-Boeck, Chen & Davison 2017).

iv. Choose your data and methods

Whatever your research question is, eventually you'll need to consider which data source and analytical strategy are most likely to provide the answers you're seeking. One starting point is to consider whether your question would be best addressed by qualitative data (such as interviews, observations or historical records), quantitative data (such as surveys or census records) or some combination of both. Your ideas about data sources will, in turn, suggest options for analytical methods. You might need to collect your own data, or you might find everything you need readily available in an existing dataset someone else has created. A great place to start is with a research librarian: university libraries always have them and, at public universities, those librarians can

work with the public, including people who aren't affiliated with the university. If you don't happen to have a public university and its library close at hand, an ordinary public library can still be a good place to start: the librarians are often well versed in accessing data sources that might be relevant to your study, such as the census, or historical archives, or the Survey of Consumer Finances.

Because your task at this point is to plan research, rather than conduct it, the purpose of this step is not to commit you irrevocably to a course of action. Instead, your goal here is to think through a feasible approach to answering your research question. You'll need to find out, for example, whether the data you want exist; if not, do you have a realistic chance of gathering the data yourself, or would it be better to modify your research question? In terms of analysis, would your strategy require you to apply statistical methods? If so, do you have those skills? If not, do you have time to learn them, or money to hire a research assistant to run the analysis for you?

Please be aware that qualitative methods in particular are not the casual undertaking they might appear to be. Many people make the mistake of thinking that only quantitative data and methods are scientific and systematic, while qualitative methods are just a fancy way of saying: 'I talked to some people, read some old newspapers, and drew my own conclusions.' Nothing could be further from the truth. In the final section of this guide, you'll find some links to resources that will provide more insight on standards and procedures governing qualitative research, but suffice it to say: there are rules about what constitutes legitimate evidence and valid analytical procedure for qualitative data, just as there are for quantitative data.

As you work through these four steps in planning your project, it's perfectly normal to circle back and revise. Research planning is rarely a linear process. It's also common for new and unexpected avenues to suggest themselves. As the sociologist Thorstein Veblen wrote in 1908 : 'The outcome of any serious research can only be to make two questions grow where only one grew before.' That's as true of research planning as it is of a completed project (Zhaojun, Paul, & Jian, 2020).

Self-Assessment Exercises 2

1. Researchers face limits in terms of … and ….?
2. Why do we review previous research?
3. State three important items to be included in a quantitative field studies.

2.5 Summary

Research planning and execution is always an iterative process. While there is no one 'best' way to design research, planning a project involves four general steps: orienting yourself to knowledge-creation; defining your research question; reviewing previous research on your question; and then selecting and analysing relevant data to formulate your own answers. After completing this unit, you'll be able to do the following: describe clearly and concisely the question you've chosen to study; summarise the state of the art in knowledge about the question, and where your project could contribute new insight and identify the best strategy for gathering and analysing relevant data.

2.6 Glossary

ANOVA: Analysis of Variance

ANCOVA: Analysis of Covariance

Execution: The execution phase is the period of active data collection and analysis.

2.7 References/Further Readings

Brooke, H. **(2022).** How to plan a research project**.** America: Aeon Media Group Limited.

Eboh, E.C. (2009). Social and economic research: Principles and methods (2nd Edition). Enugu: African Institute, for Applied Economics.

Gujarati, D.N. (2003). Basic Econometrics. New Delhi: Tata McGraw-Hill Publishing

De-Boeck, P., Chen, H., and Davison, M. (2017). Spontaneous and imposed speed of cognitive test responses. *British Journal of Mathematical and Statistical Psychology*. 70(2):225–237.

Zhaojun, L., Paul, D. B., and Jian, L. (2020). Does planning help for execution. Accessed on 31st August 2022 from https://osf.io/8pw3d/.

2.8 Possible Answers to SAEs

Answers to SAEs 1

1. The execution phase is the period of active data collection and analysis

2. Research planning involves broadly similar steps, including: formulating a question, developing an argument or predictions based on previous research, then selecting the information needed to answer your question.

3. Ethical clearance is usually a requirement for publishing findings in peer reviewed

journals.

4. In engaging staff in thesis support work the contract should specify:

- The nature of the work to be done
- The work hours
- Deliverables
- Supervision and quality assurance procedures (if necessary)
- Compensation
- Travel Allowances (if necessary)

Answers to SAEs 2

1. Researchers face limits in terms of time and money

2. The purpose of reviewing previous research is to describe the state of the art in knowledge on the research question that a project has posed.

3. For quantitative field studies, this will have included: participant inclusion criteria, exclusion criteria and sample size.

MODULE 4 RESULTS AND DATA ANALYSIS

Unit 1: Reduction and Organization of Data

Unit Structure

1.1 Introduction

In this unit, you will learn how to present data in different forms the will make it easier for us to reduce and organize data more conveniently. After studying this unit, you are expected to have achieved the learning outcomes listed below.

1.2 Learning Outcomes

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

- Discuss the meaning of results and data collection;
- Analyse the steps in data reduction and organization; and
- Evaluate the statistical tools for data analysis.

1.3 Reduction and Organization of Data

1.3.1 Meaning of Results and Data Collection

Data do not become useful until they have been analysed. Data analysis is the next step that proceeds data collection and presentation. This typically involves applying systematic and verifiable techniques, formular, and procedures to compute and derive measures of variables and their relationships. The analysis of data is the breaking down of recorded research information into meaningful components for critical examination, extraction of otherwise hidden information and making of interpretation and inferences about the relationships existing among the parts. Data analysis involves ordering, categorizing, manipulating and summarizing data, the aim of which is to reduce it in such a way as to render them useful in answering research questions and testing stated hypotheses. Data analysis is carried out by the use of statistical tools which involves the manipulation of numerical values. The nature and purpose of the study determine the types of analysis that are applicable for each research. Results and data analysis in research begins with data reduction and organization.

1.3.2 Reduction and Organisation of Data

Results processing and analysis of data involves a number of closely related operations which are performed with the purpose of summarizing the collected data and organizing these in a manner that they answer the research questions and objectives. Data reduction is the process of getting the raw data obtained from the field ready for analysis and the calculation and summarization of the descriptive statistics. Data reduction involves the following steps:

a. Field controls,

b. Editing,

c. Coding,

d. Tabulation,

e. Creation of new variables, and

f. Calculation of summarizing statistics.

a. Field controls are devices and approaches that are aimed at ensuring that field assistants measure all the elements of the sample accurately, correctly and properly record their individual characteristics such as name, address, age, qualification and other profiles in a sample book. This is to enable the researcher make necessary verifications of information obtained from the respondents when the need arises.

b. Data editing entails critical examination of the turned-in research instrument (questionnaire) in order to ensure that the information required is supplied, readable, and accurate. It is also defined as the process relating to the review and adjustment of collected survey data with an aim to control the quality of the collected data. The accuracy of the supplied data can be improved upon by checking the internal consistency of the responses. Data editing can be performed manually, with the assistance of a computer or using both. Data editing is crucial as it helps in take full advantage of the available data to be converted into useful data, ensuring that the errors arising during collection, entry, assimilation are omitted or minimized. It also assures that the consistency is coherent and consistent, since such characteristics have a constructive impact on the final analysis and outcomes (Uttarakhand 2013).

c. Coding is the categorization of the obtained responses and the assignment of data or frequency values to them. The purpose of data coding is to bring out the essence and meaning of the data that

has been collected from the respondents. In order to make sense of the data, it must be analyzed. Analysis begins with the labeling of data as to its source, how it was collected, the information it contains. Coding facilitates the researcher to reduce the bulk of information and data to a form that is easily understandable and can be interpreted soon either manually or through software programming. Coding, as an essential part of editing is done in order to get the data ready for tabulation.

d. Tabulation involves the transfer of data from the coding sheet, test score sheet, or pre-coded questionnaire or interview forms and putting them in tables. It is the process of summarizing raw data and displaying the same in compact form for further analysis. It is an orderly arrangement of data in columns and rows. To be simple and understandable, every each table should concentrate on a few ideas, attributes or variables. Besides, every table must bear a title and a number at its top. Tabulation is essential because:

- It conserves space and reduces explanatory and descriptive statement to a minimum.
- It facilitates the process of comparison.
- It facilitates the summation of items and the detection of errors and omissions.
- It provides the basis for various statistical computations.

Tabulation may also be classified as simple and complex tabulation. Simple tabulation generally results in one-way tables which supply answers to questions about one characteristic of data only. Complex tabulation usually results on two-way tables that give information about two interrelated characteristics of data, three –way tables or still higher order tables known as manifold tables. The components of data tables include:

- Table Number
- Title

- Head notes
- Stubs
- Caption
- Body or field
- Footnotes
- Source

- **Table Number**: Each table should have a specific table number for ease of access and locating. This number can be readily mentioned anywhere which serves as a reference and leads us directly to the data mentioned in that particular table.
- **Title:** A table must contain a title that clearly tells the readers about the data it contains, time period of study, place of study and the nature of classification of data.
- **Head notes:** A headnote further aids in the purpose of a title and displays more information about the table. Generally, headnotes present the units of data in brackets at the end of a table title.
- **Stubs:** These are titles of the rows in a table. Thus a stub display information about the data contained in a particular row.
- **Caption:** A caption is the title of a column in the data table. In fact, it is a counterpart if a stub and indicates the information contained in a column.
- **Body or field:** The body of a table is the content of a table in its entirety. Each item in a body is known as a _cell‘.
- **Footnotes:** Footnotes are rarely used. In effect, they supplement the title of a table if required.

• **Source:** When using data obtained from a secondary source, this source has to be mentioned below the footnote. Most often added to source is year of data collection/study. Eg. Source: Field survey, 2021.

e. Creation of new variables by a researcher is possible after tabulation and careful examination of the obtained data. For instance, if an employee supplies information on his length of service and position. The researcher might come up with new information on the rate at which employee is promoted by his employer. The use of computers greatly facilitates the derivation of new variables during data analysis in research.

f. Calculation of summarizing statistics comes after tabulating the data for each variable and showing the frequency distributions. Examples of descriptive statistics often involved in calculation and summarizing statistics include: percentage and frequency distribution, measures of central tendency/location (mean, mode, median); and measures of variability/dispersion/ spread (range, mean deviation, variance, standard deviation).

Self-Assessment Exercises 1

1. Data analysis is the next step that proceeds … and …?
2. Results and data analysis in research begins with…..?
3. What is data reduction?
4. Coding means…?
5. Outline the 4 importance of tabulation.

1.6 Summary

This unit discussed the useful processes for data reduction and organization in a research undertaking. Data reduction involves the following steps: field controls, editing, coding, tabulation, creation of new variables, and calculation of summarizing statistics which have implications for efficiency in the research process.

1.7 References/Further Readings

Ayandike, R.N.C. (2009). *Statistical Methods for Social and Environmental Sciences.* Ibadan: Spectrum Books Limited.

Brink, D. (2010). *Essentials of Statistics,* David Brink and Ventures Publishing, Aps. www.bookboon.com. Accessed 5th June, 2012.

Eboh, E. C. (2009). *Social and Economic Research: Principles and Methods.* African Institute of Applied Economics, Enugu.

Esheya, S.E. (2012). Expressive Mathematics for schools and colleges. Abakaliki : Legro Press Nig. Limited.

Fernandes, M. (2009). *Statistics for Business and Economics.* Marcelo Fernandes and Ventus Publishing APS, Aps. www.bookboon.com. Accessed 5th June, 2012.

Gupta, C.B. & Gupta, V. (2010). *An Introduction to Statistical Methods* (23rd Revised Edition). VIKAS Publishing House PVT Limited, New Delhi.

Spiegel, M.R. & Stephens, L.J. (1999). *Schaum's Outline of Theories and Problems of Statistics.* USA: McGraw Hill.

Uttarakahand (2013). Researching for Hospitality and Tourism Management. Online lecture notes of Uttarakhand Open University. Unpublished.

1.8 Possible Answers to SAEs

Answers to SAEs 1

1. Data analysis is the next step that proceeds data collection and presentation.

2. Results and data analysis in research begins with data reduction and organization.

3. Data reduction is the process of getting the raw data obtained from the field ready for analysis and the calculation and summarization of the descriptive statistics.

4. **Coding** is the categorization of the obtained responses and the assignment of data or frequency values to them.

5. Tabulation is essential because it:

- conserves space and reduces explanatory and descriptive statement to a minimum.
- facilitates the process of comparison.
- facilitates the summation of items and the detection of errors and omissions.
- provides the basis for various statistical computations.

Unit 2: Statistical Tools for Data Analysis

Unit Structure

2.1 Introduction

Data analysis is the next step that follows in every typical research project after data have been collected, organized and presented accordingly. Statistics is concerned with scientific method of collecting, organising, summarising, presenting and analysing masses of numerical data so as to comprehend the essential features and relationship of the data. As an indispensable part of the research process, statistical tools are broadly divided into descriptive and inferential statistics. In this unit we shall be considering the application of descriptive statistical tools for the analysis of research data.

2.2 Learning Outcomes

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

- Discuss the descriptive statistics used for data analysis;
- Analyse statistical information on graphs; and
- Interpret research results based on analysed data.

2.3 Statistical Tools for Data Analysis

2.3.1 Descriptive Statistical Tools

Descriptive statistics "consist of methods for organising and summarising information using tables, charts (histograms, bar charts, pie charts) and graphs. These enables the researcher to describe observed events, explain their order of importance, state the extent two sets of values or scores are related and the relativity of individual performances. Examples of descriptive statistics include percentage and frequency distribution, measures of central tendency/location (mean, mode, median); measures of variability/dispersion/ spread (range, mean deviation, variance, standard deviation); measures of relationships and measures of relative understanding.

a. **Percentages**

The percentage of a given score or number is the product of that score and 100. It is expressed as a ratio in which the second number is arranged to be 100. Translating frequency counts into percentages shows the number per hundred compared using a common base of 100 for comparison. The following point are noteworthy while using percentages for data analysis.All percentages add up to 100 because percentage is expressed on the hundred-point scale.Percentages cannot be used for the comparison of two unequal groups.Percentages should be stated in one category. For example, if 70% of the students passed the first semester examination on HTM403, it may not be necessary to state that 30% of them failed.

Example 1. The table below shows students' enrolment in various departments of the Faculty of Agricultural Sciences, NOUN. Express this information in percentages.

Table I: Distribution of students according to Departments in FAS, NOUN.

S/N	Name of Department	Number of Students
1.	Agric. Economics & Extension	214
2.	Animal Science and Fisheries	73

3.	Crop & Soil Science	46

Solution:

Table Ib: Distribution of students according to Departments

S/N	Name of Department	Number of Students
1.	Agric. Economics & Extension	214
2.	Animal Science and Fisheries	73
3.	Crop & Soil Science	46
	Total	**333**

% age for each department = $\frac{\text{No. in respective departments X 100}}{\text{Total number of students}}$

Department of Agric. Econs and Extension = $\frac{214 \text{ X } 100}{333}$ **= 64.3%**

Department of Animal Science and Fisheries = $\frac{73 \text{ X } 100}{333}$ **= 21.9%**

Department of Crop and Soil Science = $\frac{46 \text{ X } 100}{333}$ **= 13.8%**

These can be expressed in percentages as shown in the table below.

Table Ic: Distribution of students according to Departments

S/N	Name of Department	Number of Students	Percentage
1.	Agric. Economics & Extension	214	64.3
2.	Animal Science and Fisheries	73	21.9
3.	Crop & Soil Science	46	13.8
	Total	**333**	**100**

Example 2. In hotel presidential, 35 out of the 81 guests that booked for a conference checked in. What percentage of the guests were absent at the conference?
Solution:

Total number of expected guests = 81

No. of guests present = 35
No. of guests absent = 81-35 = 46

% age of guests absent = $\frac{46 \times 100}{81}$ = **56.8%**

b. Frequency Distribution

At this juncture, it is necessary to describe the concepts of ungrouped and grouped data.

i. Frequency Distribution for Ungrouped Data

In the presentation of data, data need to be arranged in an array. **An array** is the arrangement of data (observations of a variable) either in an ascending or descending order of magnitude. Data is presented as either ungrouped or grouped data. Data organised and summarised in a frequency distribution are called **grouped data** while those that have not been organised in a frequency distribution are called **ungrouped data**. The number of times a score appears is called **frequency** while the table which shows each score and the number of times it occurs is known as **frequency distribution table.**

For example, consider the score of 20 students in GST102 TMA at 10 percent.

5, 4, 7, 2, 8, 5, 5, 9, 5, 6, 8, 4, 5, 3, 10, 7, 9, 6, 5, 2.

These scores are presented in frequency distribution table as shown below which enables us to easily analyse and comprehend data. Data in this table is ungrouped because it does not specify group or class and as such cannot be convenient in handling large data.

Table 2: Frequency distribution table for GST102 scores.

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

ii.Frequency Distribution for Grouped Data

Data grouping becomes necessary when the range scores in the distribution is very large. Frequency distribution for grouped data is applicable when the data or scores obtained are very large and far apart. In this case, it becomes necessary to put the data in classes or groups to determine how many scores are in each class or category. When masses of raw data are summarised, it is important to distribute the data into categories called classes. In determining the number of individuals belonging to each class or frequency, a class width is usually developed. A tabular representation of data by classes including the corresponding frequencies is referred to as **frequency distribution.** Frequency distribution is made from an array so as to condense the data by reducing the number of rows such that closely related values are grouped onto class intervals. Consider the scores of 50 students in HTM403 first semester examination presented below:

9, 44, 48, 50, 47, 53, 56, 57, 58, 91,

90, 77, 69, 63, 18, 20, 81, 88, 83, 87,

7, 11, 23, 26, 28, 25, 75, 74, 8, 70,

79, 6, 5, 30, 37, 32, 38, 93, 12, 16,

4, 66, 64, 67, 61, 68, 62, 70, 39, 40.

Table 3: Frequency Distribution of Scores of 50 Students in HTM403 Examination

Class interval	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71 - 80	81-90	91-100
Frequency	6	5	4	6	4	4	9	5	4	3

- ***Class interval and class limits***

In Table 3, 1- 10 in the first column is called class interval. The class intervals made up the class limits (lower and upper class limits). For example, in the class interval of 1-10, 1 is the lower class limit while 10 is the upper class limit. The class limits are the smallest (lower class limit) and the largest (upper class limit) observation in each class (Esheya, 2012).

- ***Class boundaries and class marks***

Class boundaries are obtained by adding the upper limit of one class interval to the lower limit of the next higher class interval and dividing by 2. Just as we have the upper and lower class limits, we also have lower and upper class boundaries. For example, in the second column of table 3 above, the lower class limit is 11 and the upper class limit is 20. The lower class boundary is the upper class limit of the class just before it plus the lower class limit of that class and dividing by 2 as follow: $\frac{11+20}{2} = \frac{31}{2} = 15.5$

The upper class boundary of column 2 of table is the upper limit of that class plus the lower limit of the next class limit, that is: $\frac{20+21}{2} = \frac{41}{2} = 20.5$

The class boundary for the class interval 11-20 in table 3 above is 15.5-20.5 and so on.

• **Class Mark:** Class mark is the mid- point of the class interval obtained by adding up the lower and upper limits of each class and dividing by 2. For example, the class mark for column 1 is:

$\frac{1+10}{2} = \frac{11}{2} = 5.5$ and the class mark for column 2 is: $\frac{11+20}{2} = \frac{31}{2} = 15.5$

• **Class Width:** The width or size of a class is the difference between one lower class limit and the next lower class limit or difference between one upper class limit and the next upper class limit. From table 3, the width of the class = 10 – 1 or 20 – 11, which is = 9.

• ***Cumulative frequency and relative frequency***

Cumulative frequency is the number of observations in that class plus all the observations above it. It indicates the total frequency of a set of class intervals. Relative frequency is also known as percentage. It is used for comparing the frequencies in each class relative to the total frequency and multiply by 100. It is the frequency of the class divided by the total frequency of all classes and is expressed as a percentage. From table 3, the class boundary, class mark, frequency and relative frequency can be generated as shown in table 4.

Table 4: Cumulative Frequency and Relative Frequency Table

Class Interval	**Class Mark (X) or mid-point**	**Frequency**	**Cumulative Frequency**	**Relative frequency**
1-10	5.5	6	6	12
11-20	15.5	5	11	10
21-30	25.5	4	15	8
31-40	35.5	6	21	12

41-50	45.5	4	25	8
51-60	55.5	4	29	8
61-70	65.5	9	38	18
71-80	75.5	5	43	10
81-90	85.5	4	47	8
91-100	95.5	3	50	6
Σ		**50**		**100**

2.4 iii. Graphical Presentation of Data

Statistical information are often represented graphically in bar charts, pie charts, histogram, frequency polygon and cumulative frequency polygon (ogive).

- **Bar Charts**

A bar chart is a graphical display of frequency distribution in bars. This is a way of summarizing a set of categorical data. It displays the data using a number of rectangles, of the same width, each of which represents a particular category. The bars can be vertical or horizontal, which are divided into sections. Each section of a bar chart corresponds in size to the frequency of the item it represents. A space is usually left between each adjacent bar so that the categories are separated. Bar charts/graphs are represented below:

Table 5: Distribution of Rice Production in Ebonyi State

Quantity (tons)	7500	4000	5000	6000	8000

Year	2017	2018	2019	2020	2021

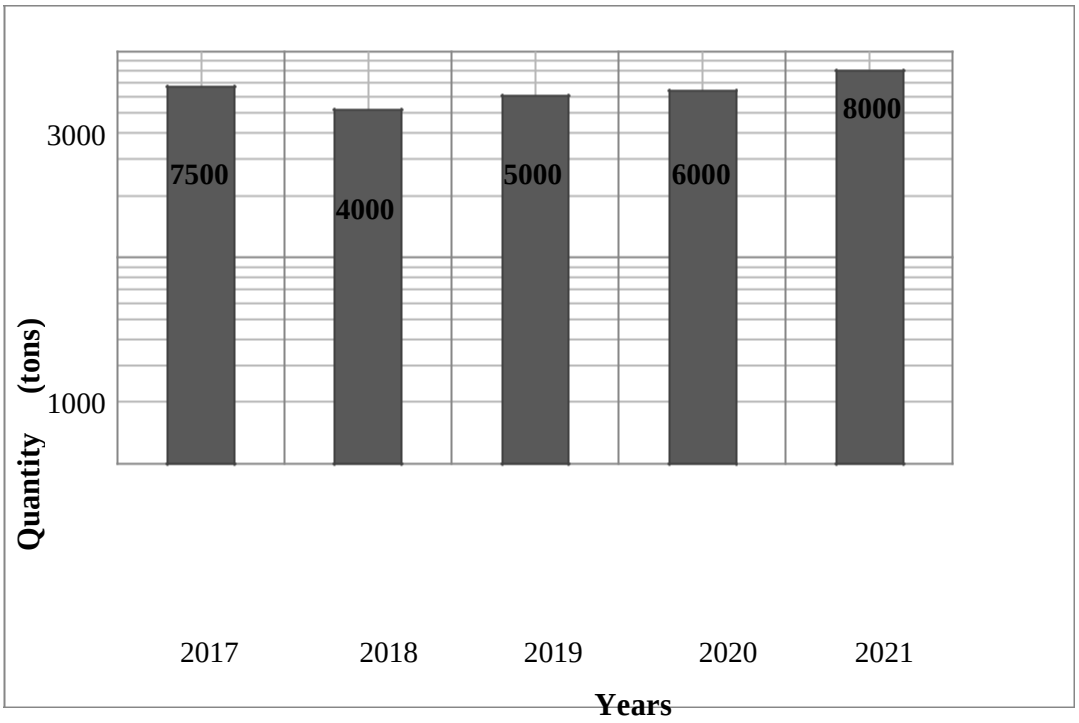


Figure 1: A bar chart showing the distribution of rice production in Ebonyi state.

- **Pie Charts**

A pie chart is a circular diagram with the circumference of 360^0, in which each item is represented by a sector whose area is proportional to the percentage or relative frequency of the total. In the drawing of a pie chart, the following steps are followed.

i. Find the angle of the sector corresponding to each item in a circumference of 360^0. The formula for finding this is $\frac{F_I}{\Sigma F_i} X\ 360$

Where: Frequency or value contributed by each item and ΣFi = the sum of all the frequencies or total value of the items.

ii. Find the relative frequency or percentage value of each item in relation to the total value of items. The formula for finding this is

$$\frac{F_I}{\Sigma F_i} X\ 100$$

For instance, the expenditure of a household on food consumption shown in table 7 below can be calculated and represented in a pie chart.

Table 6: Expenditure of a household on food consumption

Item (s)	Rice	Yam	Gari	Soup	Others
Amount (N)	74,000	43,600	50,100	61,400	82,000

Solution:

Table 6b: Expenditure of a household on food consumption

Item (s)	**Amount (N)**	**Degree of expenditure**	**Percentage of expenditure**
Rice	74,000	$\frac{74,000\ X\ 360^0}{311,100} = \mathbf{85.6^0}$	$\frac{74,000\ X100}{311,100} = \mathbf{23.8}$
Yam	43,600	$\frac{43,600\ X360^0}{311,100} = \mathbf{50.5^0}$	$\frac{43,600\ X100}{311,100} = \mathbf{14.0}$
Gari	50,100	$\frac{50,100\ X360^0}{311,100} = \mathbf{57.9^0}$	$\frac{50,100\ X100}{311,100} = \mathbf{16.1}$

Soup	61,400	$\frac{61,400 \times 360^0}{311,100} = \mathbf{71.1^0}$	$\frac{61,400 \times 100}{311,100} = \mathbf{19.7}$
Others	82,000	$\frac{82,000 \times 360^0}{311,100} = \mathbf{94.9^0}$	$\frac{82,000 \times 100}{311,100} = \mathbf{26.4}$
Total	**311,100**	**360^0**	**100**

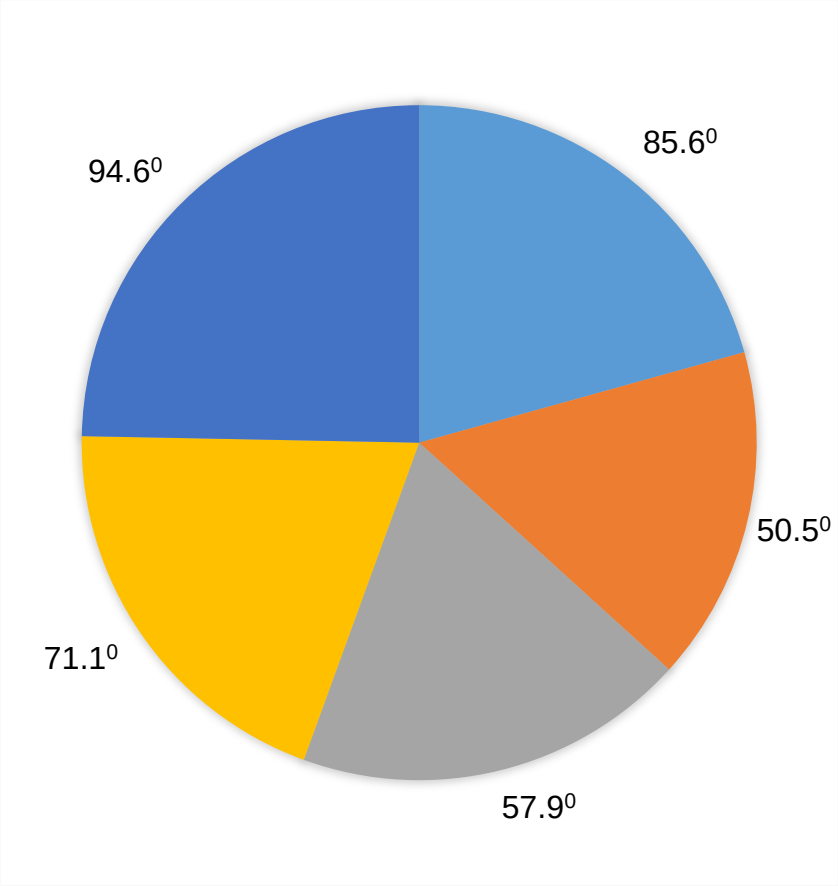


Figure 2: Pie Chart on the Household Expenditure on Food Consumption

- **Histogram**

A histogram consists of rectangles and vertical bars or columns whose heights are proportional to the frequencies in each class interval. It is often used in exploratory data analysis to illustrate the

features of the distribution of the data in a convenient form. Histograms are sometimes confused with bar charts. A histogram is used for continuous data, where the bins represent ranges of data, while a bar chart is a plot of categorical variables. The horizontal axis (X-axis) of a histogram represents the class intervals to ensure that the bars are not separated while the vertical axis (Y-axis) represents the frequency of the observations**.** An example of a histogram can be constructed from data in table 7 as shown in figure 3 below.

Table 7: Age Distribution of Staff of the University of Nigeria, Nsukka

Age (years)	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71 - 80	81-90	91-100
Frequency	0	0	2	4	4	5	3	1	0	1

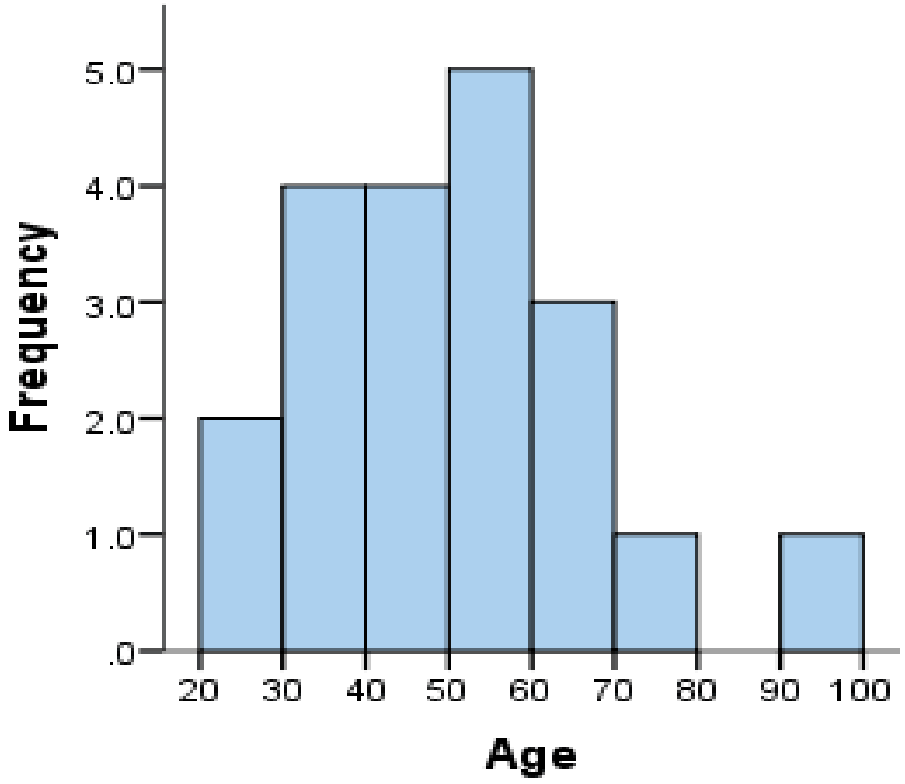


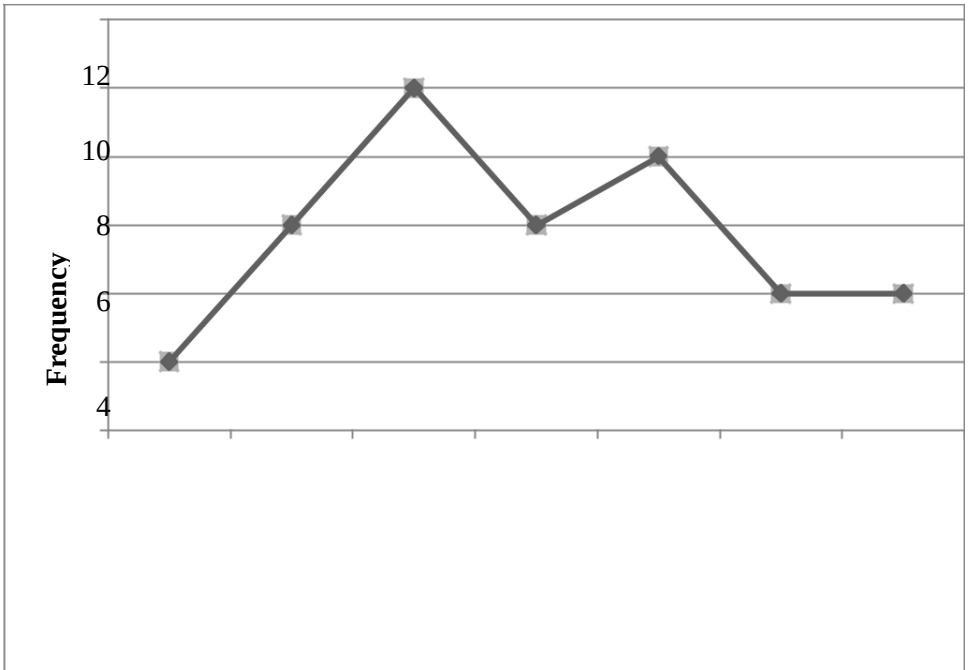
Figure 3: Histogram of ages of staff of the University of Nigeria Nsukka.

- **Frequency polygon**

A frequency polygon is drawn by joining the mid-points of the top of the histogram blocks and extending the line to both sides of the interval adjacent to it. It is a line graph joining the mid-point of the class interval. In a frequency polygon, a class mark is located in each class interval or class boundary. The horizontal axis (X axis) of the chart is made up of the class mark while the vertical axis (Y axis) is made up of frequencies. A line is normally made to join all the points together. This curve is called frequency curve or frequency polygon. In drawing the frequency polygon, you must have to construct a histogram and link the mid-points with a curve as shown below in figure 4 below.

Table 8: Population Distribution of states in Nigeria.

Population size (No./Million)	1-20	21-40	41-60	61-80	81-100	101-120	121-140
Frequency	0	0	2	4	4	5	3



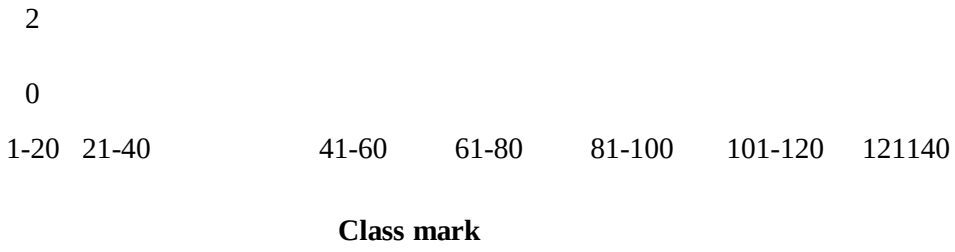


Figure 4: A frequency curve showing population distribution of states in Nigeria.

- **Cumulative frequency polygon or ogive**

For grouped data, the cumulative frequency is obtained by the addition of all frequencies from top to bottom. A cumulative frequency curve or ogive is drawn by plotting cumulative frequency against the upper limits of the class interval and joining them with a curve. A cumulative frequency curve or polygon can be formed by plotting the cumulative frequencies of the distribution on the vertical axis (Y axis) against the upper boundary on the horizontal axis (X axis). From Table 9, cumulative frequency polygon is plotted by using cumulative frequency values on the vertical axis and upper class boundary values on the horizontal axis as shown in figure 5 below.

Table 9: Distribution of Marks Scored by Postgraduate Students in an Examination

Marks (Class interval)	Frequency	Cumulative Frequency
1-10	0	0
11-20	2	2
21-30	7	9
31-40	4	13
41-50	5	18
51-60	2	20
61-70	3	23
71-80	0	23
81-90	0	23
91-100	0	23

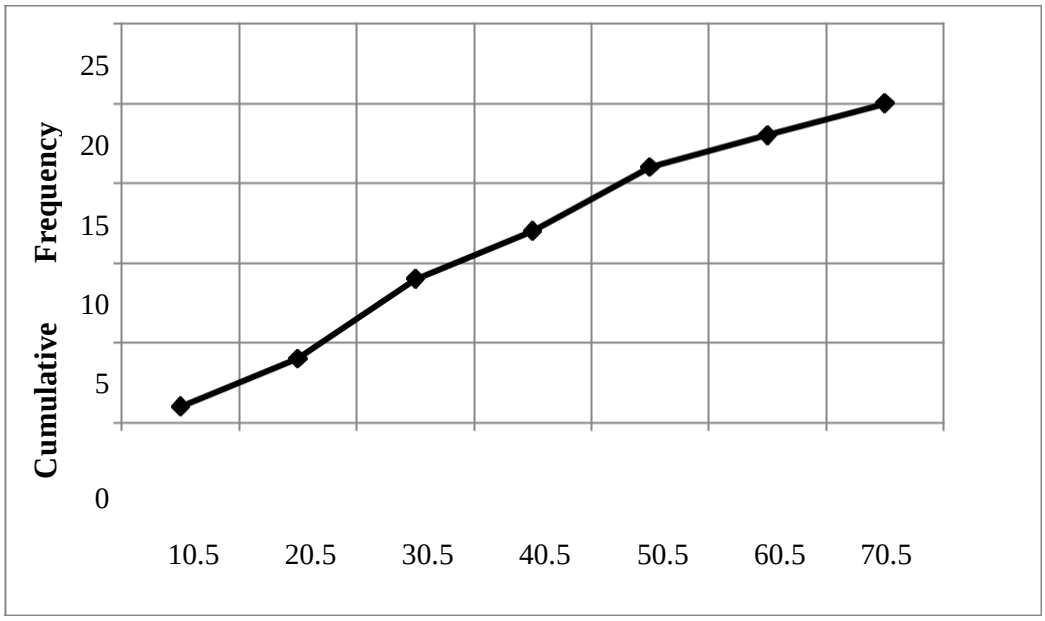


Figure 5: A cumulative frequency curve showing marks scored by Postgraduate students in an examination.

Self-Assessment Exercises 1

1. The percentage of a given score or number is ….?
2. A cumulative frequency curve or ogive is drawn by plotting cumulative frequency …..?
3. In a frequency polygon, a class mark is located in each…?
4. Differentiate between a histogram and bar chart.
5. Class boundaries are obtained by…….?

2.5iv. Measures of central tendency or location

Measures of central tendency are measures of the location of the middle or the center of a distribution. Measures of central tendency can be described as a number or value that is representative of other numbers or values by bringing the set of data towards the centre. The popular measures of central tendency are the mean or arithmetic mean, median and mode. Let us consider the mean, mode and median of ungrouped data and grouped data

.

- **The mean, mode and median of ungrouped and grouped data:**

Mean is the value or average obtained by dividing the sum of the distribution by the number of observations.it is defined as the sum of a set of scores divided by the number of scores. The symbol of mean is "X bar"($\bar{X}$).

For ungrouped data, the formular for mean $(\bar{X}) = \frac{\sum X_i}{n}$

Where:

Σ (sigma) = summation sign of.

X_i = Value of items from 1 to n.

n = Number of observations.

For grouped data, mean $(\bar{X}) = \frac{\sum fX}{\sum f}$

Where:

$\sum fX$ = Summation of the frequencies multiplied by X in each case, and $\sum f$ = Total frequency

Example 3a: Find the mean of the following scores: 21, 40, 63, 20, 18.

Solution:

Mean $(\bar{X}) = \frac{\sum X_i}{n}$ = 21+40+63+20+18

5

= 162

5 = 32.4

Example 3b. Find the mean of the scores shown in the table below.

Table 10: Mean scores of students

Score (X)	15	29	18	12	22
Frequency	3	2	5	6	4

Table 10b: Solution:

Score (x)	15	29	18	12	22
Frequency (f)	3	2	5	6	4
Fx	**45**	**58**	**90**	**72**	**88**

Mean $(\bar{X}) = \frac{\sum fX}{\sum f}$ = 45+58+90+72+88

3+2+5+6+4

= 353

20 **= 17.7**

Example 3c: Find the mean age using the information provided in the table below.

Table 11: Mean age of informants

Age (years)	1-10	11-20	21-30	31-40	41-50

Frequency	9	6	5	3	1

Table 11b: Solution:

Class interval	**Class mark or mid-point (x)**	**Frequency (f)**	**Fx**
1-10	5.5	9	49.5
11-20	15.5	6	93
21-30	25.5	5	127.5
31-40	35.5	3	106.5
41-50	45.5	1	45.5
Σ		**24**	**422**

Mean $(\overline{X}) = \frac{\Sigma fx}{\Sigma f}$ = $\frac{422}{24}$ = **17.6**

The mode is a value occurring most frequently. It is rarely of any practical use for numerical data. Figure that appears most often in the obtained data of interest. In an ungrouped data, the mode of a distribution is the value that occurs most frequently; but in a grouped data, the modal class is the class interval which has the highest frequency. Given the distribution:

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

Here 4 is the mode because it occurred 6 times which is the highest frequency in this distribution.

Example 4: Find the modal class in the distribution below.

Table 12: Modal class of distribution

Age (years)	1-10	11-20	21-30	31-40	41-50
Frequency	3	7	2	1	4

Solution:
Looking at this grouped data, the modal class is the interval between 11-20 years. This is because it has the highest frequency value of 7 in the distribution.

The median is the middle value of a distribution arranged in ascending or descending order of magnitude. The median of a series is the value that divides the series into two equal halves when arranged in ascending or descending order. The median (also referred to as the 50th percentile) is the middle value in a sample of ordered values. Half the values are above the median and half are below the median. It is the middle value of data when ranked. For ungrouped data, the median can easily be determined if the observations are arranged either in ascending or descending order of magnitude from the class interval corresponding to *N/2* of the frequency (Esheya, 2012).

Example 5: Find the median of 12, 18, 10, 15, 20, 11, 16, 17, 21, and 14.

Solution:
Rearrange the numbers in ascending or descending order as:

10, 11, 12, 14, 15, 16, 17, 18, 20, 21.

$$\text{Median} = \frac{15+16}{2} = \mathbf{15.5}$$

For grouped data, the median of the classes can also be determined. There are two methods of determining the median score from grouped data which which is outside the scope of this course.

• Relationship between mean, mode and median

When we have a distribution that has uniform number of high and low scores, the distribution is said to be **normal**. This means that the population mean, population median

and the mode are all located at the centre. This can be represented by a **symmetrical** bell-shaped curve in which the curve can be divided into two equal halves (one half is a mirror image of the other). **Skewness** refers to the degree to which the frequency distribution of a group of scores or data is asymmetrical. Asymmetrical means lack of symmetry. In mathematics, a figure is called symmetric if there exists a point in it through which if a perpendicular is drawn on the X-axis, it divides the figure into two congruent parts i.e. identical in all respect or one part can be superimposed on the other i.e mirror images of each other. In Statistics, a distribution is called symmetric if mean, median and mode coincide. Otherwise, the distribution becomes asymmetric which is not normal.

.

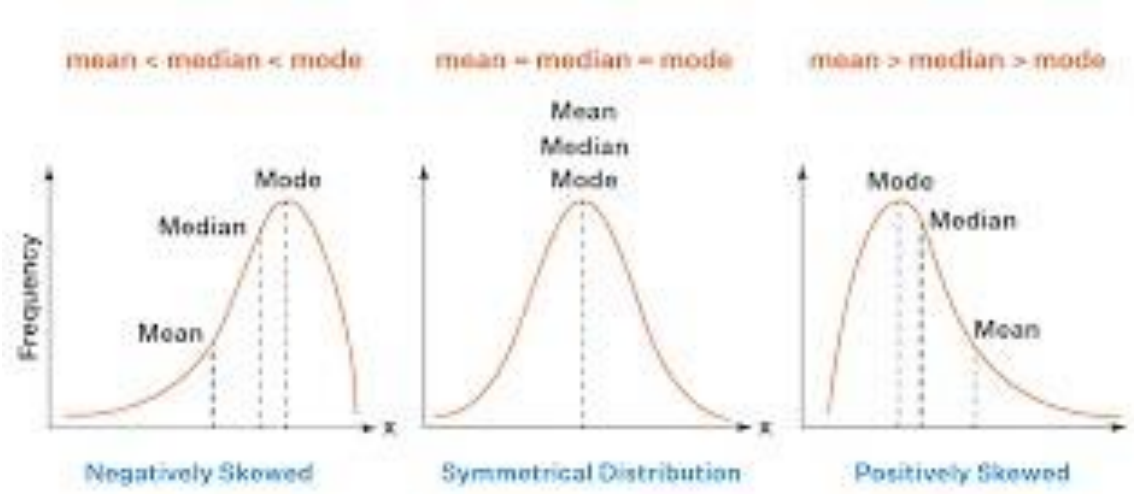


Figure 6: Distribution curve showing the relationship between mean, mode and median

When the distribution has either larger number of relatively low or high scores, then the mean, median and mode have different values and the shape of the curve is **asymmetrical.** If the distribution is such a way that the curve is negatively skewed to the left, it means that

the mean is the lowest while the mode is the highest. This implies that the set of observation has larger number of relatively large scores where mode is relatively high.

However, if the distribution has larger number of relatively low scores, then the mean has the largest value while the mode has the least. Such a situation is positively skewed to the right. This means that Mean > Median > Mode. In whichever the direction of skewness of the curve, the median always divide the curve into two (Brink, 2010)).

The mean is the best measure of central tendency although it is affected by extreme values. However, it is still reliable. The median is not affected by extreme values while the mode is easy to determine but may not be a unique value. The mean, mode and median coincide in normal distribution. The distribution is positively skewed if the mode is less than the median and median less than the mean. Negatively skewed is implied when the mode is greater than the median and median greater than the mean (Ominyi & Odo, 2006).The mean, median and mode are similar when the distribution is symmetrical. When the distribution is skewed the median is more appropriate as a measure of central tendency.

2.6v. Measures of Dispersion, Spread or Variability

A measure of variability is an index used to study the spread of data around its mean value in a distribution. This is a numerical value describing the amount of variability present in a data set. . It is used to observe the degree of spread or unevenness of a set of observation. The most common measures of dispersion are range, mean deviation, variance, standard deviation, and coefficient of variation.

• **The range** is the difference between the highest and lowest scores or values in a distribution and hence the simplest measure of dispersion. Range only takes into consideration the extreme values in a set of data. It is an unreliable measure of spread because the degree and pattern of variability often differ from one distribution to another.

Range (R) = $X_{max.} - X_{min.}$
Where:

$X_{max.}$ = highest value in the distribution
X_{min} = lowest value in the distribution.

Example 6: Find the range of: 3, 6, 8, 12, 18 and 24.

Solution: The range is 24 – 3 = **21.**

• **The mean deviation** is the arithmetic mean of the deviations from the mean, in which all the deviations are treated as positive regardless of signs. This measures the difference between the observed values and their arithmetic mean. The mean deviation of a frequency distribution is the mean of the absolute values of the deviation from some measures of central tendency such as mean. Mean deviation is obtained by ignoring the signs of the deviation and simply regarding all of them as positive. The mean of these absolute deviations is referred to as mean deviation. Steps in calculating mean deviation include:

i.Calculate the mean of the deviation that is, $\bar{X}$

ii.Record the deviation from the mean in each observation that is $d = (X_i - \overline{X})$.

iii.Convert the deviation to absolute value

iv.Find the average value of deviations (mean deviation)

For grouped data, mean deviation $= \frac{\sum(X-\bar{X})}{n}$

Example 7: Calculate the mean deviation of the distribution shown in this table.

Table 13: Mean deviation of distribution

Class interval	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71 - 80	81-90	91-100
Frequency	6	5	4	6	4	4	9	5	4	3

Table 13b: Solution:

Class interval	**Class mark (x)**	**Frequency**	**fx**	$\mathbf{d = (X_i - \bar{X})}$	$\boldsymbol{\lvert d \rvert}$	$\boldsymbol{\lvert Fd \rvert}$
1-10	5.5	6	33	5.5-48.5	43	258
11-20	15.5	5	77.5	15.5-48.5	33	165
21-30	25.5	4	102	25.5-48.5	23	92
31-40	35.5	6	213	35.5-48.5	13	78
41-50	45.5	4	182	45.5-48.5	3	12
51-60	55.5	4	222	55.5-48.5	7	28
61-70	65.5	9	589.5	65.5-48.5	17	153
71-80	75.5	5	377.5	75.5-48.5	27	135
81-90	85.5	4	342	85.5-48.5	37	148
91-100	95.5	3	286.5	95.5-48.5	47	141
		Σf = 50	**Σfx =2426**		**Σ/d/=250**	**Σ/fd/ = 1210**

Mean $\bar{X} = \frac{\sum FX}{\sum F} = \frac{2426}{50} = 48.5$

$$\text{Mean Deviation} = \frac{\sum F|X_i - \bar{X}|}{\sum F} = \frac{\sum F|d|}{\sum F} = \frac{1210}{50} = 24.2$$

- **The variance** is the square of the mean deviation. The symbol of variance is (S^2).

Variance (S^2) for ungrouped data = $\frac{\Sigma(X_i-\bar{X})^2}{n}$

Variance (S^2) for grouped data = $\frac{\Sigma f(X_i-\bar{X})^2}{\Sigma f}$

Example 8: Find the variance of the scores in the above table.

Table 14: Variance of scores

Class interval	**Class mark (x)**	**Frequency**	**Fx**	**d = ($\mathbf{X_i - \bar{X}}$)**	$(X_i - \bar{X})^2$
1-10	5.5	6	33	5.5-48.5	43^2 = 1849
11-20	15.5	5	77.5	15.5-48.5	33^2= 1089
21-30	25.5	4	102	25.5-48.5	23^2= 529
31-40	35.5	6	213	35.5-48.5	13^2=169
41-50	45.5	4	182	45.5-48.5	3^2 =9
51-60	55.5	4	222	55.5-48.5	7^2=49
61-70	65.5	9	589.5	65.5-48.5	17^2=289
71-80	75.5	5	377.5	75.5-48.5	27^2=729
81-90	85.5	4	342	85.5-48.5	37^2=1369
91-100	95.5	3	286.5	95.5-48.5	47^2=**2**209
		Σf = 50			**Σ**$(X_i - \bar{X})^2$ **=8290**

Solution: First, create and compute a column for $(X_i - \bar{X})^2$ as in the above table.

Variance (S^2) for grouped data = $\frac{\Sigma(X_i-\bar{X})^2}{\Sigma f}$ = $\frac{8290}{50}$ = **165.8**

• **The standard deviation** is commonly regarded as the square root of the variance. The name standard deviation implies that it is more standard method of measuring deviations from the mean. The symbol of standard deviation is (S).

Standard deviation (S)for ungrouped data = $\sqrt{\frac{\Sigma(X_i-\bar{X})^2}{n}}$

Standard deviation (S) for grouped data = $\sqrt{\frac{\Sigma f(X_i - \bar{X})^2}{\Sigma f}}$

Steps in calculating standard deviation:

i. Calculate the mean of a set of observations.

ii. Find the deviation of each observation from the mean.

iii. Square each deviation from the mean.

iv. Find the sum of all the squared deviations.

v. Divide the sum of all the squared deviations by the number of observations less one.

vi. Divide the square root of the value obtained in 5.

Following our last example, the standard deviation is $\sqrt{165.8}$

S = **12.9**

• **The coefficient of variation (CV)** is a relative measure of dispersion expressed as percentage. Coefficient of variation is defined as the standard deviation divided by the mean and multiplied by 100.

$\boldsymbol{CV = \frac{S}{\bar{X}} X\ 100}$

Where:

CV = Coefficient of variation;

S = Standard deviation;

$\bar{X}$= Mean of the distribution.

Coefficient of variation is used in comparing relative dispersion of two or more sets of observations. It has the advantage that it standardizes the variability of distribution by their means and so allows for the differences in the average trends not accounted for by the standard deviation alone. The group with higher coefficient is said to have a wider spread.

The C.V in our last example is $\frac{12.9\ X\ 100}{48.5}$ = **26.6**

Self-Assessment Exercises 2

1. List three popular measures of central tendency.
2. Given a variance of 144, calculate its standard deviation.
3. Coefficient of variation is defined as …?
4. What is mean deviation?
5. A distribution is positively skewed if.…?

2.6 Summary

In this unit, you have learnt in details the descriptive statistical tools used in data analysis including measures of location and variability from both ungrouped and grouped data. You have also learnt the necessary steps involved in the calculation of mean, mode, median, range, mean deviation, variance, standard deviation as well as coefficient of variability. Standard deviation is also called "root mean square deviation". The variance of a set of data is defined as the square of the standard deviation **while the c**oefficient of variation is defined as the standard deviation divided by the mean and multiplied by 100.

2.7 References/Further Readings

Afonja, B. (2005). *Introductory Statistics: A Learner's Motivated Approach*. Ibadan: Evans Brothers Nigeria Publishing Limited.

Ayandike, R.N.C. (2009). *Statistical Methods for Social and Environmental Sciences.* Ibadan: Spectrum Books Limited.

Brink, D. (2010). *Essentials of Statistics*, David Brink and Ventures Publishing, Aps. www.bookboon.com. Accessed 5th June, 2012.

Esheya, S.E. (2012). Expressive Mathematics for schools and colleges. Abakaliki : Legro Press Nig. Limited.

Fernandes, M. (2009). *Statistics for Business and Economics.* Marcelo Fernandes and Ventus Publishing APS, Aps. www.bookboon.com. Accessed 5th June, 2012.

Gupta, C. B. & Gupta, V. (2010). *An Introduction to Statistical Methods* (23rd Revised Edition). New Delhi: VIKAS Publishing House PVT Limited.

2.8 Possible Answers to SAEs

Answers to SAEs 1

1. The percentage of a given score or number is the product of that score and 100

2. A cumulative frequency curve or ogive is drawn by plotting cumulative frequency against the upper limits of the class interval and joining them with a curve.

3. In a frequency polygon, a class mark is located in each class interval or class boundary.

4. A histogram is used for continuous data, where the bins represent ranges of data, while a bar chart is a plot of categorical variables.

5. Class boundaries are obtained by adding the upper limit of one class interval to the lower limit of the next higher class interval and dividing by 2.

Answers to SAEs 2

1. The popular measures of central tendency are the mean or arithmetic mean, median and mode.

2. Its standard deviation = 12.

3. Coefficient of variation is defined as the standard deviation divided by the mean and multiplied by 100.

4. This measures the difference between the observed values and their arithmetic mean.

5. The distribution is positively skewed if the mode is less than the median and median less than the mean

Unit 3: INFERENTIAL STATISTICAL TOOLS

Unit Structure

3.1 Introduction

In statistical analysis, it has been observed that a change in one variable can bring about changes in another variable or variables if they are related, dependent, associated or correlated. Inferential statistics is a tool for establishing the degree of relationship or association between two or more variables. This provides the means for evaluating relationships that exist within the data collected from a sample of the population. In this unit, we shall consider some important inferential statistical tools such as Correlation coefficient, Regression analysis, Chi-square, T-test, Z-test, F-Test, Analysis of Variance (ANOVA) and Analysis of Covariance (ANCOVA).

3.2 Learning Outcomes

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

- Discuss the meaning of inferential statistics;
- Demonstrate examples of inferential statistical tools; and
- Evaluate the application of these tools to data analysis and interpretation.

3.3 Inferential Statistical Tools

3.3.1 Definition of Inferential Statistical Tools

As defined earlier, inferential statistics which is also called the parametric data analysis is a statistical tool for establishing the degree of association or relationship between two or more variables. These tools enable the researcher to test formulated hypotheses and make generalization or inferences to the population a particular sample was drawn from. Inferential statistics are more rigorous than descriptive statistics in their application to data analysis because it deals with inferences about populations based on the behavior of samples. Inferential statistics

is concerned with gaining knowledge about a population's characteristics from information collected from a random sample of the population. This implies that the inferential statistical data obtained from observation on the sample are generalized to the population from which the sample was drawn. This enables researchers to undertake more complex economic, social and behavioral problems. Some important inferential statistical tools to be considered in this context are Correlation coefficient, Regression analysis, Chi-square, T-test, Z-test, Analysis of Variance (ANOVA), and Analysis of Covariance (ANCOVA).

3.3.2 Correlation Coefficient

Correlation is a measure of the degree of relationship existing between two variables. It could be defined as the degree to which variables are related hence, it is an index for measuring the degree to which variables are associated. The measure of this relationship is known as correlation coefficient (r). Correlation measures the strength and direction of relationship between two or more variables. It is a measure of the degree of association between dependent and independent variables. The dependent variable (Y) is the one whose behaviour is been investigated as been influenced by the independent variables (Xs). For example, if X increase as Y increases, there is positive correlation. However, if Y decreases as X increases, there is negative or inverse correlation. If there is no apparent direction of movement between X and Y, then there is no correlation or X and Y are uncorrelated.

If there is no apparent direction of movement between X and Y, then there is no correlation or X and Y are uncorrelated. If all the points on scattered diagram line up directly, then there is perfect correlation between X and Y. Simple correlation involves one dependent and one independent variables; while multiple correlation involves one dependent and two or more independent variables. The precise magnitude of correlation between X and Y can be determined using correlation coefficients called Karl Pearson Correlation or Product Moment Correlation (r). The value of r ranges from -1 to +1.

Types/ Forms of Correlation

i. Linear correlation

Correlation is considered linear when all points on a scattered diagram seem to cluster near a straight line. However, a relationship could either be positive or negative. Two variables are said to be linearly and positively correlated if they tend to move in the same direction as can be observed in the supply function: **Qs= f(P).** In this function, an increase in the price of the commodity leads to an increase in the quantity of the commodity supplied, *ceteris paribus.*

Besides, two variables are said to be negatively and linearly correlated if the two variables tend to move in the opposite direction. Thus, if one of the variables is increasing the other variable will be decreasing showing an inverse relationship between the two variables. This type of relationship is shown by the demand function: **Qd = f(P).** The demand theory states that the higher the price of a commodity, the lower will be the quantity demanded.

Therefore, there is an inverse relationship existing between price, P and quantity demanded, Qd.

ii. Non-linear Correlation

Correlation between two variables is said to be non- linear when all points seem to lie near a curve. As in linear correlation, non-linear correlation could be positive or negative. The correlation between two variables, X and Y is said to be positive non-linear if they move in the same direction described by a curve. Also, two variables are negatively non-linearly correlated when they move in opposite directions (that is) an inverse relationship along a curve.

iii. Zero Correlation

This type of correlation occurs when two variables are uncorrelated with each other. In other words, there is no relationship between the two variables. On the scattered diagram, the points are dispersed all over the surface of the plane with no suitable line to join them hence, the term zero correlation.

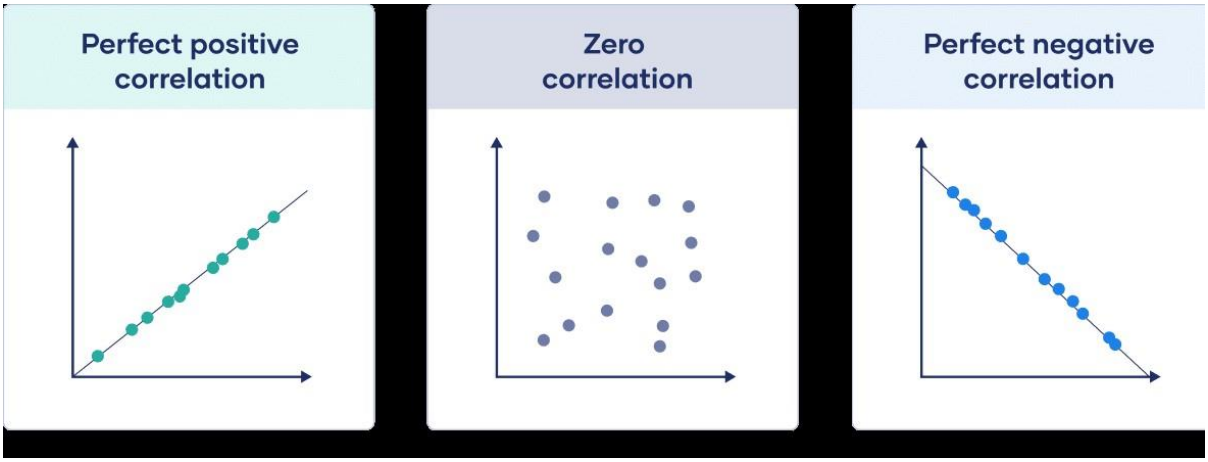


Figure 7: Types of Correlation: Perfect positive correlation, Zero correlation, and Perfect negative correlation

Computing Correlation Coefficient Using Rank Correlation Coefficient (Spearman's Correlation Coefficient)

The magnitude of the correlation existing between variables is termed the correlation coefficient. This coefficient measures the degree of co-variability of two variables, X and Y. The correlation coefficient has a range from +1 to -1 i.e. $-1 \leq r \leq 1$. To calculate this correlation coefficient, two approaches could be taken.

The Rank Correlation Coefficient (Spearman's Correlation Coefficient) requires that the observations be ranked in a specific sequence. After assigning ranks to the observations, we can then measure the relationship between the observations' ranks instead of their actual numerical values. If two variables are ranked, their rank (Spearman's) correlation coefficient can be calculated from the formula:

$$r^{t} = 1 - \frac{6\sum D^{2}}{n(n^{2}-1)}$$

Where:

D = difference between ranks of corresponding pairs of X and Y.

n = Number of observations

The values of r' has a range of $-1 \leq r' \geq 1$.

In applying this method of correlation analysis, it is necessary to note that:

(i). it does not matter whether the observations are ranked in ascending or descending order. However, the same rule of ranking must be used for both variables; and

(ii). if two or more observations have the same value, we assign to them rank.

Example 1: The following table shows the ranking preference of two customers for four (4) different brands of cement: Unicem, Bua, Dangote, and Elephant. Calculate the rank correlation coefficient, r.

Table 15: Ranking preferences of two consumers

Brands of cement	Unicem (U)	Bua (B)	Dangote (D)	Elephant (E)
Ranking of customer X	2	4	5	4
Ranking of customer Y	3	1	2	6

Solution:

The differences between the rankings of the two customers is given by:

Difference, D = Ranking of customer X and Y

The difference, D is shown below:

Table 15b: Solution

Brands of cement	U	B	D	E

Ranking of customer X	2	4	5	4
Ranking of customer Y	3	1	2	6
Difference D	-1	3	3	-2
D^2	1	9	9	4
ΣD^2	1+9+9+4 = **23**			

Substituting values in the formula:

$$r^{t} = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

$$r = 1 - \frac{6 \text{ X } 23}{4\,(4\text{X}4-1)} = 1 - \frac{138}{60} = 1 - 2.3 = \mathbf{-1.3}$$

Therefore, the rank correlation coefficient, r = -1.3 and this shows a fairly different preferences of the two customers (X and Y) for the four brands of cement under consideration. Note that the existing relationships or correlation is negative and also strong.

Interpretation of Correlation Coefficient

The correlation coefficient obtained from this method is a measure of degree of relationship between two variables. A correlation coefficient can take any value from -1 to +1, and can be interpreted as follows:

Table 16: Interpretation of Correlation Coefficient Value

Correlation Coefficient Value	**Interpretation of Correlation for Positive Values**	**Interpretation of Correlation for Negative Values**
1.00	Perfect positive	Perfect negative
0.8 to 0.99	Very high positive	Very high negative

0.6 to 0.79	Moderate positive	Moderate negative
0.4 to 0.59	High positive	High negative
0.2 to 0.39	Low positive	Low negative
0.1 to 0.19	Very low positive	Very low negative
0	No or zero correlation	

Self-Assessment Exercises 1

1. Inferential statistics is a tool for….?
2. What is correlation coefficient?
3. Correlation is considered linear when…?
4. Inferential statistics is also called …?
5. Zero correlation means …..?

3.4 Regression Analysis

The study of most disciplines today, including hospitality and tourism management is based on establishing relationship between variables. Regression analysis is a tool which describes in mathematical form, the relationship between variables. In statistics, regression analysis is a technique which examines the relation of a

dependent variable to specified independent variables. This provides for measuring the magnitude of relationship between variables. The key relationship in a regression is the regression equation which contains regression parameters whose values are estimated using data. In other words, regression analysis presents an equation for estimating the amount of change in the value of one variable associated with a unit change in the value of another variable. For instance, given the regression model: $Y = \beta_0 + \beta_1 X + \mu$. Y is referred to as the **predictor** or **dependent** variable, X is the **explanatory or independent** variable, and β_0 and β_1 are **coefficients.** It is implicitly assumed that the explanatory variable X "causes" the dependent variable Y, and the coefficient β_1 measures the influence of X on Y.

Types of regression analysis

Regression analysis could be simple, multiple, linear or non-linear.

i. **Simple Regression**

This is a regression analysis which describes in mathematical form, the relationship between two variables. It is also called the ***two-variable linear regression model*** or ***bivariate linear regression model*** because it relates two variables. This implies that the relationship has one dependent variable and only one independent variable. Suppose we wish to estimate the parameters of our reference demand function, stated in the implicit form $\{Q_d = f(P)$, we can express it in the explicit form as: $Q_d = \beta_0 - \beta_1 P + \mu$ and employ the technique of the simple regression analysis to estimate its parameters; β_0 and β_1.

The explicit form above implies that there is a one-way causation between the variables Q_d and P. Price, P is the cause of change in the quantity demanded, but not vice versa. Hence we talk of regressing quantity demanded of the commodity "Q_d" against the price "P".

Where:

Q_d = Quantity demanded of the commodity;

P = Price of the commodity;

β_0 = intercept (constant); and

β_1 = the slope coefficient.

The variable *u*, called the **error term** or **disturbance** in the relationship, represents factors other than P that affect Q_d such as taste, presence of other substitutes, wars etc. A simple regression analysis effectively treats all factors affecting Q_d other than P as being unobserved. You can usefully think of *u* as standing for "unobserved."

ii. Multiple Regression analysis

This is a regression that involves the relationship with more than two variables. Therefore, the multiple regression analysis is applied to a model with one dependent (predictor) variable. Hence, any model with a minimum of two independent variables requires multiple regression technique for its analysis. For example, the quantity demanded of any commodity (Q_d) depends on such factors as Price (P) of the commodity, Price (P*) of its close substitute, and consumer's income (Y) among others. The relationship between the quantity demanded and these factors can be written in its implicit form as:

$$Q_d = f(P, P^*, Y).$$

Explicitly, the above demand function can be written as:

$$Q_d = \beta_0 - \beta_1 P + \beta_2 P^* + \beta_3 Y + \mu$$

Considering the above equation, it can be seen that Q_d is the dependent variable, while the independent variables are P, P*, and Y. To estimate the parameters (β_0, β_1, β_2, and β_3) of this demand model, we require the application of the multiple

regression analysis (or technique). It is important to note that the number of independent variables relative to the existing coefficients can be extended to nth number as the case may be.

i. Linear or non-linear regression

Regression analysis may be linear or non-linear. Linear regression emanates from linear equations in the sense that we can fit a straight line to a set of data involving the dependent and independent variables. The consumption function: C = f(I) is an example of a linear equation which can be transformed to a linear regression model as: $C = a_o + \mathbf{bI} + \mu$. Non- linear regression models develop of quadratic and cubic equations. A popular example of a non-linear regression model is the Cobb Douglas Production Function: Y = f(L, L*, C,M).

Computation of regression parameters: regression equation

This method is considered the best way or technique of obtaining the line of best fit. In using the Ordinary Least Square (OLS) method to fit regression line, the following procedures have to be followed. **First,** assume a linear relationship between the dependent variable, Y and (independent) variable, X. Hence, express the relationship thus: Y = a + bX

Where a = intercept

b = coefficient of X.

Example 2: Given Y and X as shown below:

Y	69	76	52	56	57
X	9	12	6	10	9

Solution:

Table 17: Compute the data for OLS model of Y and X

No. of observations (n)	**X**	**Y**	**XY**	**X^2**
1	9	69	(69x9) = 621	9x9 = 81
2	12	76	(76x12) = 912	12x12 = 144
3	6	52	(52x6) = 312	6x6 = 36
4	10	56	(56x10) =560	10x10 = 100
5	9	57	(57x9) = 513	9x9 = 81
n = 5	**ΣX = 46**	**ΣY= 310**	**ΣXY= 2918**	**ΣX^2 = 442**

The regression line to be estimated is $\hat{Y} = \hat{a} + \hat{b}X$. Here, that is in regression analysis, the hat or cap (^) means estimated.

$$a = \frac{\sum X^2 \sum Y - \sum X \sum XY}{n\sum X^2 - (\sum X)^2};$$

$$\hat{b} = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$$

Note:
n is the number of observation, in this case there are 5 observations, that is n = 5.

ΣY = 310: This means summation (addition) of all the numbers under Y.

ΣX = 46: This means summation (addition) of all the numbers under X.

ΣXY = 2918: This means summation (addition) of all the numbers under XY.

ΣX^2 =442: Taking the square of each value of X to get X^2 of each and then sum them up.

To calculate the intercept a **,** substitute the values into the formular: $a = \frac{\sum X^2 \sum Y - \sum X \sum XY}{n\sum X^2 - (\sum X)^2}$;

$= \frac{442 \text{ X } 310 - 46\text{X}2918}{5\text{X}442 - (46)^2}$

$= \frac{2792}{94}$

a **= 29.7**

To calculate the coefficient, $\hat{b}$, of X, also substitute values accordingly.

$\hat{b} = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$

$= \frac{5 \text{ X } 2918 - 46 \text{ X } 310}{5 \text{ X } 442 - (46)^2}$

$= \frac{330}{94}$

$\hat{b}$ **= 3.5**

Substituting the value of a and $\hat{b}$ in the regression model, $\hat{Y} = \hat{a} + \hat{b}X$we have:

Y = 29.7 + 3.5X

To plot the graph of the relation Y = 29.7 + 3.5X, we substitute the values of X in table 2 above into the regression model, Y = 29.7 + 3.5X, and find the corresponding values of Y.

Interpretation of parameter regression estimates

In a simple regression analysis, we estimate two things: the regression constant ($\widehat{\beta_0}$) and the regression coefficient ($\widehat{\beta_1}$).

$$\hat{Y} = \widehat{\beta_0} + \widehat{\beta_1}X + \mu_t$$

Where:

$\widehat{\beta_0}$= Regression constant, $\widehat{\beta_1}$= Regression Coefficient

Generally, the regression constant is a real value affecting the dependent variable determined by some factors outside the relationship between the dependent variable and the independent or explanatory variable. Regression constant will have physical meaning to the problem and will be the average value of Y when X is zero. This is only possible if these two conditions hold:

i. it must be physically possible for X to equal zero; and
ii. data must be collected around X =0

When the two conditions do not hold, the regression constant is then considered as a real value affecting the dependent variable determined by something outside the relationship between the dependent and independent variable.

Besides, the regression coefficient ($\widehat{\beta_1}$) is used to observe existing difference and not changes overtime. Rather than talk about a change, we need to talk about two things that are, at a given point in time, different from each other.

Consider the estimated regression equation below:

$$\hat{Y} = 3.7 + 0.9X$$

Where:

$\hat{Y}$ = Estimated maize yield (in kg)

X = Amount of fertilizer applied

We can interpret the above regression equation thus:

The regression constant $\widehat{\beta_0}$=3.7 is the average maize yield when no fertilizer is applied (i.e. when X = 0). This interpretation is possible since some quantity of maize will be harvested when no amount of fertilizer is applied to the plot.

On the other hand, the regression coefficient, $\widehat{\beta_1}$ = 0.9 denotes that if two plots are treated with different quantities of the same fertilizer such that the quantity of fertilizer applied to the two different plots differ by 1 unit, then on the average, the maize yield in the plot with higher quantity of fertilizer will be 0.9kg more than the maize yield in the plot with smaller quantity of fertilizer.

3.5 Chi-Square Test

Chi-square (χ^2) is a statistical technique which enables the researcher to evaluate the probability of obtaining differences between the actual and expected frequencies in the categories of one or more classifications as a result of sampling fluctuation. It is a non-parametric statistics used to determine the frequency or proportion with which events or objects occur. Simply put, Chi-square Is a measure of the discrepancy between observed and expected frequencies. The Chi-square (χ^2) test determines whether the observed frequencies of a given observation differ significantly from the frequencies which might be expected from an assumed hypothesis. The chi square is a non-parametric statistic. Generally, the result obtained in a sample does not always agree with expected result. A chi square distribution is used to determine the difference between expected and the observed results. There are basically two methods involved in chi square testing:

i. Test of goodness of fit

ii. Test of independence

The simplest form of the test for independence is found when there are only two groups within each basis of classification giving 2 X 2 contingency table; but it is possible to have more than two groups depending on the researcher's basis of classification. Here, a cross classification table is created, the basis in rows is designated 'r' and the one in the column as 'c'. The entire table is called r X c contingency table. Chi-square is applied when proportions are divided into two-folds and multi-fold. In a-two fold situation, there are only two possible answers: yes or no, male or female. Such proportions are mutually exclusive categories. Chi-square formula is:

$$\chi^2 = \Sigma \frac{(fo - fe)}{Fe}$$

Or

$$\Sigma\frac{(o - e)}{e}$$

where:

χ^2 = chi-square
Σ = summation (sigma)
fo = frequency observed
fe = frequency expected.

Conditions and Steps for the Application of (χ^2) Test

The necessary conditions for applying Chi-square (χ^2) test are as follow:

i. Observations recorded and used are collected on a random basis.

ii. All the items in the sample must be independent.

iii. No group should contain very few items, say less than 10. In case where the frequencies are less than 10, regrouping is done by combining the frequencies of adjoining groups so that the new frequencies become greater than 10.

iv. The overall number of items must also be reasonably large. It should normally be at least 50, however small number of groups may be accepted.

v. The constraints must be linear. Constraints which involve linear equations in the cell frequencies of a contingency table (i.e., equations containing no squares or higher powers of the frequencies) are known are know as linear constraints.

Properties of Chi-Square Distribution

i.Chi-square (χ^2) is skewed to the right.

ii.The distribution is concentrated on the right hand side of the curve (no negative value).

iii.The exact shape of the distribution depends on the degree of freedom i.e. there are different chi square distribution curve for different degrees of freedom.

iv.The distribution tends to shift to the right and become flatter for large values of the degrees of freedom.

v.The degrees of freedom (V) is usually the number of observations (sample size) minus 1 (k – 1) for test of goodness of fit and (r -1) (c -1) for test of independence, where contingency table is used.

Note: r = number of rows and c = number of columns.

Example 3: Consider the null hypothesis Ho: There is no significance difference between public and private schools in Nigeria. If a sample of 90 respondents are randomly selected from different states in the country, using chi-square analysis, you can determine numerically those who are in support of public and those for private schools.

Solution:
Assuming from observation, 60 said 'Yes' to public school and 30 said 'No'.

Table 18: Frequency of responses

Answers	**Observed frequency**	**Expected frequency**
Yes	60	45
No	30	45
	90	**90**

When two cases are involved, degree of freedom (Df) = 2-1 =1. Then 0.5 must be subtracted

from the absolute difference between the observed and expected.

Table 18b: Analysis of responses

Answers	**Observed**	**Expected**	**/O –E/ - 0.5**	**/O – E/-0.5)^2**	**/O – E/-0.5)^2 / E**
Yes	60	40	+19.5	380.25	9.5
No.	30	60	+19.5	380.25	6.3
					X^2 = 15.8

Chi-square is = χ^2

Calculated chi-square = χ^2

Critical/hypothetical chi-square = χ^2e

Df = k-1 2-1 =1

At 0.5 significant level and df = 1, the hypothetical or critical chi-square X^2e = 6.314 at 5% probability. In this case, our calculated chi-square is 15.8, which is greater than X^2e of 6.314.

Based on this result, the null hypothesis **Ho**: There is no significance difference between public and private schools in Nigeria, is rejected. This implies the acceptance of the alternative hypothesis **H_A**: There is significance difference between public and private schools in Nigeria (χ^2 is Chi-Square). $x^2 \neq \chi^2$

Self-Assessment Exercises 2

1. Given the regression model: $Y = \beta_0 + \beta_1 X + \mu$, define the parameters
2. Chi-square is applied when …..?
3. In a simple regression analysis, we estimate two things: … and …?
4. Simple regression means …..?
5. What is a linear constraint?

3.6a. T-test (Student's test)

If a sample of size n is taken from a normal population, the variable t can then be defined as:

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\bar{X} - \mu\sqrt{n}}{S}$$

Where:

S = the sample standard deviation

The formula is analogous to the Z-statistic given by:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}} = \frac{\bar{X} - \mu\sqrt{N}}{\sigma}$$

The sample standard deviation is:

$$S = \sqrt{\frac{\sum(X - \bar{X})^2}{n - 1}}$$

The t distribution is usually used for small sample (n≤30). We find it convenient to replace n with n – 1 and we now have the sample standard deviation (S_x)

$$S_x = {}^{S}/_{\sqrt{n-1}}$$

Instead of having Z distribution, we have t distribution

$$t = \frac{\bar{X} - \mu}{{}^{S}/_{\sqrt{n-1}}}$$

Where:

n -1 = the number of degree of freedom (V). The first person that brought about this was W. S. Gosset. He was an employee of Guinness Brewery in Dublin, Island. Guinness Brewery was the first industry to employ statistician and they kept it secret. They did not want to use his name Gosset, rather, they called it studen's t distribution after its discoverer who published his work in the name „student" during the early twentieth century.

The concept of the degree of freedom (V) is used in different areas of statistical analysis. Essentially, the number of degree of freedom indicates the number of values that are free to vary in a random sample. In general, the number of degree of freedom lost is equal to the number of population parameters estimated as the basis of statistical inference. Therefore, the degree of freedom = n -1, where n is the sample size.

Example 4: Suppose we have a sample that contains five items and we know that the mean is 20, then the sum should be 100.

Solution: So, we are free to assign arbitrary value to the rest of 4 items while the last one is fixed. 100 – (10 + 25 + 30 + 18) = 83. It means that the 5^{th} item must be 17. So, if we fix the sample mean at X in a sample of n, we can assign values to any variable up to n – 1 item in the sample. Thus, the name degree of freedom = n-1.

Testing for differences in means

In testing for differences in means, there is the need to develop the null hypothesis, i.e. There is no difference in the means. Then the t distribution is given by

$$= \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^{\,2}}{n_1} + \frac{S_2^{\,2}}{n_2}}}$$

$$= \frac{\bar{X}_1 - \bar{X}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

For $(n_1 + n_2)$ degree of freedom, the value of S is obtained as:

$$S = \sqrt{\frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}}$$

Example 5: The monthly sale of a provision seller was 120 cartons. After an increase in the demand for provisions, the mean monthly sale increased to 125 cartons for 20 kiosks

and a standard deviation of 15 cartons. Show your views whether the seller is making good sales or not at 5% level of significance.

Solution:

H_0 = the seller is not making good sale

$$\bar{X} = 125, \mu = 120, and\ \sigma = 15$$

$$t = \frac{\bar{X} - \mu\sqrt{n}}{S} = \frac{125 - 120\sqrt{19}}{15} = \frac{5\ (4.36)}{15} = \frac{21.8}{15}$$

t-cal. = 1.45

The number of degree of freedom = n – 1 = 19.

At 5% level of significance, **the critical t-value = 1.73**

Decision: The t-calculated value is less than the critical value (1.45 < 1.73) at 5% level of probability. Therefore the H_0 is accepted and we conclude that the seller is not making good sale.

3.6b. The Z-Distribution

As a result of central limit theory, the normal distribution (Z distribution) can be used as a basis for determining the critical value for testing a hypothesis. Z is usually employed when n ≥ 30. The formula for Z is:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}} = \frac{\bar{X} - \mu\sqrt{N}}{\sigma}$$

In testing for difference in means, the Z distribution can be computed as:

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Example 6: An ADP zonal manager claims that the mean monthly income of Village Extension Agents (VEAs) is ₦1400. A student wishes to test this claim by contacting 36 VEAs. The monthly income of ₦1400 is given the benefit of doubt. Test whether this claim is true at 5% level of probability, given the following information below:

$H_{0:}\ \mu$ = N1400 σ = 50, $\bar{X} = N1250$, Z-critical value - 1.96.

Solution:

H_0: X = μ (1250 = 1400)

$$Z = \frac{\bar{X} - \mu}{\sigma} = \frac{1250 - 1400}{50} = \frac{-150}{50} = -3$$

Decision/Conclusion: Hence the calculated Z value is less than the critical value (1.96) at 5% level of probability, we accept the null hypothesis that there is no significant difference between the population mean income and the sample mean income.

3.6c. The F -Distribution

The F- distribution is used to compare the significant difference in several means at the same time for the analysis of variance (ANOVA). F- test is called analysis of variance. ANOVA is a method of reducing to a ratio the proportion of the total variance due to variation between the means of the samples and the proportion due to the variation within the samples as a result of sample error.

F ratio: F ratio is the ratio of the variation between sample means and the variation within the samples.

$$F = \frac{Variation\ between\ sample\ means}{\text{Varation within sample means}}$$

The degree of freedom from this ratio is used to develop the F distribution curve. There are some similarities and differences between t distribution and F distribution.

Table 19: Differences between t Distribution and F Distribution

s/no.	**t-Distribution**	**F Distribution**
1.	Has one degree of freedom	Has two degrees of freedom, one for numerator and the other for the Denominator
2.	Has different t curves for different degrees of freedom	Has different F curves for different Pairs of degree of freedom
3.	The t- distribution curves are symmetrical	The F distribution curves are skewed to the right
4.	The t distribution curves is centred at zero	All scores in F distribution are positive
5.	The t distribution can be used for testing two means at a time	The F distribution can be used for testing several means at a time
6.	The t distribution can be used for testing both 1 and 2 tail test of hypothesis	The F distribution is used for testing 1 tail test of hypothesis

The name F distribution was named after R. A. Fisher. The F distribution comprises of two degrees of freedom V_1 and V_2. The V1 is the numerator which is = k-1, where K is the number of samples. V2 is the denominator = k (n -1), where n = sample size. n – k if n≠k.

$$F = \frac{Variation\ between\ sample\ means}{Variation\ within\ the\ samples} \frac{MSA}{MSE}$$

Where:

MSA = Mean square of A

MSE = Mean square due to error

MSA= $\frac{SSA}{K-1}$

MSE= $\frac{SSE}{K(n-1)}$

Where:

SSE = Sum square of error

k = number of samples

n = Sample size.

Self-Assessment Exercises 3

1. The name F distribution was named after….?
2. The F- distribution is used to compare…?
3. Z is usually employed when …?
4. Write the formula of t-test.
5. In testing for differences in means, there is the need to develop ….?

3.7 Analysis of Variance (ANOVA)

Analysis of Variance (ANOVA) is a statistical tool used to detect differences between experimental group means. ANOVA is a statistical test for detecting differences in group means when there is one parametric dependent variable and one or more independent (categorical) variables. ANOVA is a statistical technique used for apportioning the variation in an observed data into its different sources. In other words, it is a technique employed to break down the total variation in an experiment to its additive components.

With this technique or method therefore, we can break down the total variation, occurring in a dependent variable, into various separate factors causing the variation. ANOVA technique can be seen as a technique which can be used to assign the total variation in an experiment to the individual factors used in the experiment, combination of factors and /or to nature (i.e. chance).

In adopting this technique for the analysis of data, assumptions concerning the nature of statistical relationship between the dependent and the independent variables are not considered or made. Rather, certain assumptions are made concerning the treatments of the experiments. Thus, ANOVA does not stipulate any functional relationship between the dependent and the independent variable but it enables us to break down the total variation into the different sources or causes of the variation. However, to understand this statistical technique clearly, certain associated terms/terminologies need some clarifications. These include:

a. Factors: They are independent variables to be studied in an experiment. They are variables whose effect on another variable to be studied. Example is the study of the effect of carbohydrate on the weight gain in chicken. The factor here is carbohydrate.

b. Factor Levels: They are groups within each factor. They are sources of a particular factor. Hence, in the study of the effect of carbohydrate on the weight gain in chicken, the factor levels are the different sources of the factor (carbohydrate) under study. These sources include; wheat, dried cassava peels, maize and others. Each of these sources is a factor level.

c. Treatment: A treatment is equivalent to a factor level in a single-factor analysis. However, in multi-factor analysis, a treatment is equivalent to a combination of factor levels. For instance, in the study of the effect of carbohydrate on the weight gain in chicken, each of the factor levels (that is wheat, dried cassava peels and maize) is a treatment under a single-factor analysis.

In considering treatment as a combination of factor levels, consider this illustration. Suppose we wish to study the effect of different breeds of chicken and different sources of carbohydrates on the weight gain in chicken. We may decide to use three (3) different breeds of chicken and three (3) sources of carbohydrate, and this will give a total of nine (3 x 3 = 9) treatments. This is so because each of the sources of carbohydrate will combine with each breed of chicken to form a treatment. This combination of factor levels can be illustrated thus:

Sources of carbohydrate (FA):

1. Dried Cassava peels (C)
2. Wheat (W)
3. Maize (M)

Breeds of Chicken (FB):

1. Harco (H)
2. Nera (Hypeco)
3. Amo Sanders (S)

FB FA	C	W	M
H	HC	HW	HM
P	PC	PW	PM
S	SC	SW	SM

The treatments in this experiment are: HC, HW, HM, PC, PW, PM, SC, SW and SM

d. ANOVA Table: It is a table that shows in summary form, the computations for analysis of variance (ANOVA). This table enables us to have a quick and convenient assessment of the sources of variation, their respective sum of squares and degrees of freedom which are results of ANOVA.

Importance of ANOVA

Analysis of variance is a very important analytical tool developed by R.A. Fisher for the analysis of experimental data. Hence, it is employed in such fields of study as agriculture, medicine, engineering, economic and other social researches to analyze data and achieve logical conclusions. Besides determining the various factors which cause variation of the dependent variable, ANOVA could also be useful in the following areas:

(i) testing the overall significance of the regression;

(ii) testing the significance of the improvement in fit obtained by the introduction of additional explanatory variables in the function;

(iii) testing the equality of coefficients obtained from different samples;

(iv) testing the stability of coefficients of regression; and

(v) testing restrictions imposed on coefficient of a function.

Basically, analysis of variance technique is used to test hypothesis concerning population means; and it is employed in regression analysis to conduct various tests of significance. Hence, it is an essential component of any regression analysis.

Types of analysis of variance

Before adopting this analytical technique, we consider the number of criteria (factors) to be studied. Based on this, ANOVA has two basic classifications or types, namely one- way or one- factor ANOVA and two-way or two-factor ANOVA.

i. One- Factor ANOVA

In one-factor analysis of variance, there are only two variables: one dependent variable and one independent variable. It is used to find out the effect(s) of the single independent variable (factor) on the dependent variable. For instance, in the study of the level of maize yield using different sources of phosphorus, we are considering the effect of phosphorus on the yield of maize. Hence, phosphorus is the independent variable (factor), while the yield of maize is dependent on it. Based on this, we adopt the one-factor analysis of variance to determine if the variation in maize yield is due to the treatment (phosphorus application) or due to chance. However, to adopt and apply this one-factor ANOVA, the following assumptions must be made:

i. The treatment effect is fixed, that is T_r is fixed.

ii. The total effect of the treatment is equal to zero; that is

iii. The sum expected value of the effect is equal to zero i.e,

iv. The error is normally and independently distributed with mean zero and variance,

Consider the following yield equation;

$$Y = \mu + T_r + e_i \quad \text{----------(i)}$$

Where:

Y = Yield of the crop

μ = Population mean

ii. Two-Factor ANOVA

Unlike the one-factor ANOVA, the two-factor classification contains more than two variables. It is made up of one dependent variable and two independent variables (factors). It is basically employed to determine the effects of the two independent variables (factors) on the dependent variable. This technique, therefore, enables us to estimate not only the separate effects of the factors (independent variables) but also the joint effect (interaction effect) of these factors on the dependent variable.

Consider the study of the effects of fertilizer and soil type on the yield of cassava. There are two independent variables or factors namely; fertilizer and soil type, and one dependent variable (cassava yield). Hence, the effects of these two factors (fertilizer and soil type) can be analyzed using the two factor analysis of variance technique. By using this technique, the separate effect of the two factors on the yield of cassava will be determined. Also, the combined effect (interaction effect) of the two factors on cassava yield can be shown. However, when the two factors do not have any interaction effect, we say that they have additive effects on the dependent variable (cassava yield). In applying the two-factor

analysis of variance technique, the same assumptions holding in a one-factor technique still hold.

iii. ANOVA Table

This is a table which presents, in summary form, the computational results of the analysis of variance. It shows the values for the degree of freedom (df); sum of squares (SS), Mean squares (MS) and F-ratio. This table is in two forms; one-factor ANOVA table and two-factor ANOVA table.

One-Factor ANOVA Table

This summarizes the results for the degree of freedom (df), sum of squares (SS) and mean squares (MS) for the total variation, treatment variation and variation due to error for one-factor experiment. It equally presents the F- ratio value which is used for hypothesis testing. The following is the summary of the ANOVA table for a one-factor experiment.

Table 20: ANOVA Table for One-Factor Experiment

Source of Variation	**Sum of Squares (SS)**	**Degree of Freedom (DF)**	**Mean Squares (MS)**	**F-ratio**
Total	$\sum_{jk=1}^{N} Y_{jk}^2 - CT$	N – 1	$\frac{\sum_{jk=1}^{N} Y_{jk}^2 - CT}{N-1}$	$\frac{Mean\ Square\ treat}{Mean\ Square\ err}$
Treatment (between treatments)	$\sum_{j=1}^{t} Y_j^2 - CT$	t -1	$\frac{\sum_{j=1}^{t} Y_j^2 - CT}{t-1}$	
Residual (Error) (within treatments)	SS Total – SS Treatments	N – t	$\frac{SS\ Total - SS\ Treatm}{N-t}$	

Where:

CT = Correlation terms = $\frac{\left(\sum_{jk=1}^{N} Y_{jk}\right)^2}{N}$

r = Number of replications for treatment

t = Number of treatments

N = Total number of observations

$\sum_{jk=1}^{N} Y_{jk}^2$ = sum of the squares of all the observation in the experiment.

$\sum_{j=1}^{t} Y_{j}^2$ = sum of the squares of all the observations per treatment.

$\sum_{jk=1}^{N} Y_{jk}$ = sum of all the observations in the experiment.

Two-Factor ANOVA Table

In this case, the table presents in the summary, the results for the degree of freedom sum of squares and mean squares for the total variation, treatment variation of the two factors, and variation due to randomness (error). Equally, the F-ratio values are evaluated and presented in this table. The summary of a Two-Factor ANOVA table is shown below.

Table 21: ANOVA Table for Two- Factor Experiment

Source of Variation	**Sum of Squares (SS)**	**Degree of Freedom (DF)**	**Mean Squares (MS)**	**F-ratio**
Total	$\sum_{jk}^{N} X_{jk}^2 - C$	N – 1	$\frac{\sum_{jk}^{N} X_{jk}^2 - CT}{N-1}$	
Treatments (Between Treatments)	$\frac{\sum_{j=1}^{t} X_j^2}{b} - CT$	t -1	$\frac{SS\ \text{Treatment}}{t-1}$	$\frac{MS\ treatment}{MS\ error}$
Block (Between Blocks)	$\frac{\sum_{k-1}^{b} X_k^2}{t} - CT$	b -1	$\frac{SS\ \text{Block}}{b-1}$	$\frac{MS\ \text{Block}}{MS\ error}$
Error (Random)	SS Total – (SS Treatment + SS Block)	(t-1)(b-1)	$\frac{SS\ \text{Error}}{(t-1)(b-1)}$	

Where:

b = Number of blocks used

t = Number of treatments

$$\frac{\left(\sum_{jk=1}^{N} Y_{jk}\right)^2}{N}$$

CT = correction term

N =Total number of observations

$\sum_{jk}^{N} X_{jk}^2$ = Sum of the squares of all the observations in the experiment

$\sum_{j=1}^{t} X_j^2$ = Sum of the squares of all observations per treatment

$\sum_{k-1}^{b} X_k^2$ = Sum of the squares of all observations per block.

There are two major computational approaches for the analysis of variance (ANOVA) namely the econometric or regression and the statistical approaches which are beyond the content of this course.

3.8 Analysis of Covariance (ANCOVA)

ANCOVA provides the only viable option for analyzing data from experimental situations where randomization is not feasible. It is essentially analysis of variance in which difference between groups are tested, after controlling for other variables termed covariates. ANCOVA is sometimes regarded as ANOVA using adjusted scores. ANCOVA is most appropriate in the analysis of data

from the non-equivalent control group design. It is used to control any initial group differences, which randomization could not handle (Ominyi & Odo, 2006).

Self-Assessment Exercises 4

1. ANOVA is a statistical technique used for apportioning…?
2. ANCOVA provides the only viable option for analyzing data from…?
3. What is ANOVA table?
4. In multi-factor analysis, a treatment is equivalent to …?
5. Mention two basic classifications of ANOVA.

3.9 Summary

In this unit, we explained correlation coefficient and its interpretation. Regression analysis, its computation and the interpretation of its parameters were considered. Chi-square, t-test, z-test, and f-test as well as their applications, properties and differences were examined. We equally looked at ANOVA and ANCOVA, its meaning, terminologies, importance and types. It is observed that while analysis of variance technique provides only information concerning the breakdown of the total variation, regression analysis, in addition to this information, also provides numerical values for the influences of various independent variables on the dependent variable.

3.10 Glossary

Factors: They are independent variables to be studied in an experiment

ANOVA: Analysis of Variance

ANCOVA: Analysis of Covariance

OLS: Ordinary Least Square

3.11 References/Further Readings

Esheya, S.E. (2012). Expressive Mathematics for schools and colleges. Abakaliki : Legro Press Nig. Limited.

Gujarati, D.N., Porter, D. C. & Gunasekar, S. (2012). Basic Econometrics 5th Edition, Tata McGraw Hill Educational Private Limited, New Delhi, India.

Koop, G. (2000). Analysis of Economic Data. John Willey & Sons Ltd, New York, USA.

Koutsoyiannis, A. (2001). Theory of econometrics: an introductory exposition of econometric methods (2nd edition). Plagrave, Hampshire, New York, USA.

Ominyi, C.N. and Odo, F.A. (2006). Research and statistics in education, management and social sciences. Enugu: Mason Publishers.

Osuala, A. E. (2010). Econometrics: Theory and Problems. ToniPrints Services, Abia State, Nigeria.

3.12 Possible Answers to SAEs

Answers to SAEs 1

1. Inferential statistics is a tool for establishing the degree of relationship or association between two or more variables.

2. The magnitude of the correlation existing between variables is termed the correlation coefficient.

3. Correlation is considered linear when all points on a scattered diagram seem to cluster near a straight line.

4. Inferential statistics is also called the parametric data analysis.

5. This type of correlation occurs when two variables are uncorrelated with each other.

Answers to SAEs 2

1. Given the regression model: $Y = \beta_0 + \beta_1 X + \mu$.

Y is referred to as the **dependent** variable,

X is the **explanatory or independent** variable,

β_0 and β_1 are **coefficients and**

μ is the error term.

2. Chi-square is applied when proportions are divided into two-folds and multi-fold.

3. In a simple regression analysis, we estimate two things: the regression constant ($\widehat{\beta_0}$) and the regression coefficient ($\widehat{\beta_1}$).

4. This is a regression analysis which describes in mathematical form, the relationship between two variables.

5. A constraint which involves linear equations in the cell frequencies of a contingency table.

Answers to SAEs 3

1. The name F distribution was named after R. A. Fisher.

2. The F- distribution is used to compare the significant difference in several means at the same time for the analysis of variance (ANOVA).

3. Z is usually employed when n ≥ 30.

4. If a sample of size n is taken from a normal population, the variable t can then be defined as:

$$t = \frac{\bar{X} - \mu}{S/_n} = \frac{\bar{X} - \mu\sqrt{n}}{S}$$

5. In testing for differences in means, there is the need to develop the null hypothesis

Answers to SAEs 4

1. ANOVA is a statistical technique used for apportioning the variation in an observed data into its different sources.
2. ANCOVA provides the only viable option for analyzing data from experimental situations where randomization is not feasible.

3. This is a table which presents, in summary form, the computational results of the analysis of variance.

4. In multi-factor analysis, a treatment is equivalent to a combination of factor levels.

5. ANOVA has two basic classifications or types, namely one- way or one- factor ANOVA and two-way or two-factor ANOVA.

MODULE 5 RESEARCH REPORT AND TECHNICAL WRITING

UNIT 1 Research Proposal

UNIT 2 Research Report

UNIT 3 Application of Statistics and Computer in Research

Unit 1: Research Proposal

Unit Structure

1.1 Introduction

Research is a systematic investigation to find answers to a knowledge problem or need. Through a research proposal, the scientist communicates the essence and methodology of the contemplated study. It shows the limit to what the researcher wants to do in terms of area, scope, techniques, time frame, human and material resources application as described in this unit.

1.2 Learning Outcomes

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

- Discuss research proposal;
- Analyse the steps in preparing a research proposal; and
- Describe the components of a research proposal.

1.3 Research Proposal.

1.3.1 Meaning of Research Proposal

A proposal is an account of what an individual intends to do and how he or she wants to go about it. A research proposal is a study plan that is to be followed in the process of research efforts. The research proposal is a tool to communicate the research intent to the society. A research proposal elaborates and describes research intent. A research proposal is documentation (sequential ordering) of the research plan. The research plan is the expression of the thinking and direction of the research as conceived by the researcher at any point in time.

It is a description of the rationale, purpose, and methodology of the research. It provides a formal and detailed statement of research plan. Researchers are often required to prepare a research

proposal for submission to the Research Supervisor or Research Committee, for presentation in a Research Seminar or Workshop, for submission to Research funding agencies and networks to select for funding. Also, government and private institutions often ask for competitive research proposal on particular subjects or issues. Be that as it may, the research proposal is a document to convince project supervisors, sponsors, academic institutions, government and non-governmental organisations that the project is worthy of their attention and sponsorship. Regardless of the consumer of the research, there is always the need for a research proposal.

1.3.2 Steps in Preparing a Research Proposal

i.Choose an area of research interest

The first step is to choose an area of research interest. The area of research is the broad thematic subject of investigation, the field of study. In economics, for example, a broad thematic subject may be income distribution in a community or productivity of manufacturing enterprises.

ii.Read the literature and discuss with other researchers

Having chosen the broad thematic area of research, the next step is to embark upon extensive and intensive survey and reading of the literature on the subject. Reviewing the literature means the process whereby a researcher undertakes to search for, critically study and evaluate what has been done and what is already know about the problem.

iii. **Choose the topic and title of the research**

Following the literature study, the researcher is poised to choose the research topic.

The research topic is the specific, clear functional definition of the issue for the research. The research topic defines the specific dimension of the subject being investigated. By the end of a successful literature review, the researcher is well equipped with subject matter information, ideas and insights to identify a research topic.

Self-Assessment Exercises 1

1. Research is a systematic investigation to find answers to ……?

2. A research proposal is a ….?

3. Mention 4 areas where a research proposal may be needed.

4. Define table of content in relation to research proposal.

1.4 Components of a Research Proposal

A research proposal consists of several interacting sections. The actual ordering of the sections differs depending on the approach of individual researchers or as prescribed by the organisation for which the research is being undertaken. However, researchers in the social sciences follow generic format in writing a research proposal such as:

i.Summary/Abstract

The summary is an encapsulation of what is contained in the research proposal. Even though this section comes first in the research proposal, it is usually written last. This is because it is only when the proposal writing process is completed that the summary can be well written. The summary should contain brief and precise information on the problem or need to be addressed by the research objectives, the procedures and methods that will be used, the resource needs of the research and the likely outcomes and benefits from the research. The summary has to be written in such a manner to excite, enthuse, interest or convince the reader about the overall significance of the research. Generally, it is recommended that the summary should range from 150-350 words as the case may be.

ii.**Table of Contents**

This is a listing of all the major contents of the research proposal. It outlines the sections, subsections and gives the pages where they are found.

iii.**Introduction/Background Information**

This section is devoted to stating the significance of the research ideas. The introduction section should flow into the statement of problem section so that both sections make clear and unambiguous statement of the significance of the research.

iv. **Statement of Problem**

A research problem refers to a perceived or felt need which is capable of being resolved or ameliorated by information from a scientific enquiry. Research problem provides the setting to situate the research questions. It represents the reasons behind your proposal and specifies what is intended to be changed or improved through the research. It states the gap in knowledge to which this research intends to fill.

v. **Statement of Research Questions**

Research questions are short and precise statements of knowledge requirements to be answered by the investigation. Research questions focus on revealing the critical aspects of the research problem. The research questions constitute the heartbeat of the research problem.

vi. **Working Hypotheses**

Working hypothesis is tentative assumption made in order to draw out and test its logical or empirical manifestations. The working hypothesis delineates and focuses the research. It shapes the type of data to be collected, the analysis to be done and affect the manner in which the empirical tests will be conducted. Not all research requires a working hypothesis.

vii. **Literature review (if required as separate section)**

While some proposals integrate literature analysis in the introduction and problems statement, others have it as separate section.

viii. **Justification/Significance of the Study**

This section identifies the expected impact of the research. The Justification/significance section indicates the research outcomes in terms of the contribution to: specific problem, process of policy formulation, and societal development.

ix. **Objectives of the study**

Objectives are the measurable outcomes of the research. They refer to the tangible results of the research activity, that is, what the research will directly accomplish. The study objectives indicate what is intended to be achieved through the research. It connotes the benchmark against which the research will be adjudged to be successful.

x. **Research Design**

This section should contain descriptive of research design, sampling, measurement, and data collection/management and data analysis. The purpose of this section is to describe how the research will be carried out.

xi. **Budget**

The budget is the statement of proposed expenditures in the research. It involves the costing of activities of the research. The research budget sets out the resource implications and funding needs of the research. The budget usually includes direct and indirect expenses. Direct expenses include travel costs lodging and subsistence, documentation, materials, equipments and instruments, professional and support services. Indirect costs include institutional overheads such as rent,

utilities, salaries and wages and so on. The indirect costs are usually calculated as a percentage of the direct costs.

xii. **Work plan and Time table**

Research funding agencies require researchers to prepare work plan and time table. The work plan sets out activities, deliverables and milestones. The time table is a schedule of work according to time of implementation and completion. Sometimes, the work plan and time table are integrated in a single format.

xiii. **References**

References should be used to justify and support arguments, make comparisons between different researches and demonstrate appreciation and familiarity with the research subject. Authors should not be quoted out of context or misrepresented through indiscriminate and unjustified attributions.

xiv. Appendix (if necessary)

An appendix, where it exist, should contain additional information that is not directly part of the research proposal but which are considered to be relevant for the understanding of the intended research. Annex, such as institutional profile, curriculum vitae of researchers, data collection instrument, analytical framework, etc. should be added where applicable.

1.5 Sample of a Research Proposal

i. Cover Page

CONTRIBUTIONS OF CASSAVA PROCESSING TO HOUSEHOLD POVERTY ALLEVIATION IN SOUTH-EAST, NIGERIA.

A RESEARCH PROPOSAL SUBMITTED TO

NATIONAL OPEN UNIVERSITY OF NIGERIA

For Sponsorship

By

Dr. Samuel E. Esheya (Lead Researcher)
Dr. Peter I. Nwandu (Co-Researcher I)
Dr. Sylvanus I. Ogbonna (Co-Researcher II)

Department of Agricultural Economics and Extension
NATIONAL OPEN UNIVERSITY OF NIGERIA
June, 2022

ii.Background to the Study.

Cassava is one of the fastest expanding staple food crops with numerous uses and by products. Significant post-harvest losses of cassava roots occur in many parts of Nigeria, as a result of the tuber's inherent high moisture content, which accelerates microbial deterioration and unpredictable biochemical changes (Onyenwoke & Simonyan, 2014. Many studies on cassava in Nigeria focused on production, yet after production about 40% of cassava products are lost due to lack of power, lack of storage facilities, and lack of proper processing and packaging techniques. This reduces the quantity of cassava available for consumption therefore exacerbating food insecurity problems.

iii. Statement of the Problem

Nigeria is the largest producer of cassava in the world, but a greater percentage of cassava produce is lost or wasted by the producer due to the perishable nature of the cassava roots. Cassava roots deteriorate after three days of harvesting due to lack of storage and/or inadequate processing facilities, hence a certain percentage of cassava products is lost and/or wasted. It is imperative to assess cassava processing in the study area as this will increase rural income, employment and investment opportunities.

iv. Objectives of the Study

The main objective of this study is to determine the contributions of cassava processing to household poverty alleviation in south east Nigeria. The specific objectives are to:

i. identify the nature of losses of cassava in the study area;
ii. identify the various cassava products in the area in order of importance;
iii. identify the problems of cassava processing in the area; and
iv. ascertain the cost and returns of cassava proessing in the study area.

v. Justification of the Study

Result of this study will give signals for formulation of appropriate policies towards capacity building of the processors of African most important staple (cassava) for sustainable poverty reduction and food security. It can further be made available to scholars and policy makers all over Africa and around the world. The target audience are policy makers at all levels of government, researchers and food processors.

vi. Research Methodology

- **Study area**

The study will be carried out in South-east zone of Nigeria comprising Abia, Anambra, Ebonyi, Enugu and Imo States.

- **Sampling techniques and sample Size**

Multi-stage random sampling technique will be used in selecting the respondents. Ten agricultural zones will be selected for the cross-sectional survey of the study (two from each state). Five LGAs will be selected from each zone based on intensity of cassava processing activities. From each of the fifty LGAs, ten cassava processors will be randomly sampled bringing the number of the respondents for the cross sectional survey to five hundred (500).

- **Data collection and analytical techniques**

A well-structured questionnaire will be used for data collection and the data that will be collected will be analysed using both descriptive statistics and inferential statistics. Descriptive statistics such as frequency, percentage, mean, charts, graphs, partial budgeting, principal component factor analysis and Analysis of Variance (ANOVA) will be employed in data analysis. The test of significance in the mean ratings of the three groups of respondents (household/cottage processors, microprocessors and small/medium scale processors) will be determined using Analysis of Variance (ANOVA) at 0.05 level of significance.

vii. Expected Results

The expected results of the study include:

i. It will describe the nature of cassava product losses in the area.
ii. It will reveal the various cassava products in the area in order of importance.
iii. It will identify the problems of cassava processing in the study area.
iv. The cost and returns from of cassava processing methods employed by the processors will be ascertained.

viii. References

FAO (2008). Hunger on the Rise. *Food and Agriculture Organisation.* Retrieved April 7, 2011 from http://www.fao.org/newsroom/.

Onyenwoke, C.A. & Simonyan, K.J. (2014). Cassava post-harvest processing and storage in Nigeria: a review. *Africa Journal of Agricultural Research: 9(53), 3853–3863.*

Wong, J. & Schuchard, R. (2011). Adapting to climate change: A guide for food, beverage, and agriculture Companies. Retrieved October 5, 2021 from http://www.bsr.org/./BSR_Climate_Adaptation_Issue_Brief_Food_Bev_Ag.pdf.

Ziska, L., Runion, G. B., Tomecek, M., Prior, S. A., Torbet, H. A. & Sicher. R. (2009). An evaluation of cassava, sweet potato and field corn as potential carbohydrate sources for bioethanol production in Alabama and Maryland Biomass and Bioenergy: 33, 1503-1508.

ix. Proposed Work plan of the Study

Table 22: Proposed Work plan of the Study

Phase	*Duration*	*Activity*
1.	*4 weeks*	*Study preparation*
2.	*4 weeks*	*Framework phase of the study*
3.	*4 weeks*	*Training of enumerators*

4.	*6 weeks*	*Distribution of questionnaire*
5.	*6 weeks*	*Retrieving of questionnaire*
6.	*6 weeks*	*Data collation and coding*
7.	*6 weeks*	*Editing of data*
8.	*6 weeks*	*Analysis of data*
9.	*6 weeks*	*Report writing*
	Total	*48 weeks = 12 months.*

- ***Budget summary***

Table 23: *Budget summary*

S/N	Description	Amount (₦)
1.	Personnel cost	460,000
2.	Transportation and travels	2,100,000
3.	Training of enumerators	180,000
4.	Questionnaire Development, Production, Distribution and Retrieval	510,000
5.	Collation, Coding, Consultancy and Analysis of Data	280,000
6.	Other indirect cost	₦630,000
	Total	**4,160,000**

- **Breakdown of the budget**

Table 24: Breakdown of the budget

i. Personnel Cost

S/N	Allowances	Description	Unit Cost (₦)	Amount (₦)
1.	Principal Researcher	Principal Researcher to earn ₦20,000 per month for 10 months duration	15,000 x 10	200,000
2.	Co-Researchers	Co-Researchers to earn ₦10,000 for 10 months duration x 2	7,500 x 2 x 10	200,000

3.	Research Assistants	Three Research Assistants to be paid ₦5,000 per week for 4 weeks duration	5,000 x 3 x 4	60,000
	Sub-Total			**460,000**

ii. Transportation and Travels for Study Preparation

S/N	Allowances	Description	Unit Cost (₦)	Amount (₦)
1.	Lead Researcher	Visitation and sensitisation of the cassava processors and Local Government offices in the study area at ₦10,000 per visit to each community	10,000 x 40 communities	₦400,000
2.	Co-Researchers	Visitation and sensitisation of the cassava processors in the study area at ₦10,000 per visit to each community	10,000 x 2 x 40 communities	₦800,000
3.	Research Assistants	Visitation and sensitisation of the cassava processors in the study area at ₦3,000 per visit to each community	3,000 x 3 x 10 communities	₦900,000
	Sub-Total			**2,100,000**

iii. Training of Enumerators

S/N	Allowances	Description	Unit Cost (₦)	Amount (₦)
1.	Lead Researcher	Training of enumerators	50,000	₦50,000
2.	Co-Researchers	Training of enumerators	20,000 x 2	₦40,000
3.	Research Enumerators	Stipend to enumerators	10,000 x 3	₦30,000
3.	Research Enumerators	Entertainment of enumerators and others during training	₦60,000	₦60,000
	Sub-Total			**₦180,000**

iv. Questionnaire Development, Production, Distribution and Retrieval

S/N	Allowances	Description	Unit Cost (₦)	Amount (₦)
1.	Lead Researcher	Questionnaire development and interview schedules	50,000	₦50,000
2.	Lead Researcher	Questionnaire production	500 x 400 copies	₦200,000

3.	Lead and Co-Researchers	Administration of questionnaire and retrieval	5,000 x 400 communities	₦200,000
4.	Research Enumerators	Administration of questionnaire and retrieval	2,000 x 10 x 3	₦60,000
	Sub-Total			**₦510,000**

- **Collation, Coding, Consultancy and Analysis of Data**

Table 25: Collation, Coding, Consultancy and Analysis of Data

S/N	Allowances	Description	Unit Cost (₦)	Amount (₦)
1.	Lead Researcher	Collation of data from completed questionnaire	100 x 400 copies	₦40,000
2.	Lead and Co-Researchers	Coding of data into excel sheet for analysis	100 x 400 copies	₦40,000
3.	Data Analyst	Consultation for analysis of data from computer run-out	200,000	₦200,000
	Sub-Total			**₦280,000**

- **Other Indirect Cost**

Table 26: Other Indirect Cost

S/N	Activity	Description	Unit Cost (₦)	Amount (₦)
1.	Typing and printing	Typing and printing of final report	100 x 100 expected pages	₦10,000
2.	Journal publication	Research findings will be published in Scopus (Q1 and Q2 journals) at $500 per article	$500 x 2	₦300,000
3.	Technical Presentation	The Lead and Co-Researchers will attend an institution based national conference to present the research findings at 40,000 each	40,000 x 3	₦120,000

4.	Miscellaneous	5% of Total Cost	200,000	₦200,000
	Sub-Total			**₦630,000**
	GRAND TOTAL			**₦4,160,000**

Self-Assessment Exercises 2

1. A research problem refers to…?
2. What constitutes the heartbeat of the research problem?
3. Not all research requires a working hypothesis. True or false?
4. The significance section indicates the research outcomes in terms of the contribution to: …., ---------, and ……..?
5. List any five data analystical tools.

1.6 Summary

Preparing a research proposal is anything but a neat, continuous process with well-defined steps, each of which is completed before the next and never revisited or retraced. Instead, the whole process is interactive, that is, the researcher must continuously go "forth and back" adjusting earlier sections in the context of later segments of the proposal as described in this unit.

1.7 References/Further Readings

Eboh, E.C. (2009). Social and economic research: Principles and methods (2nd Edition). Enugu: African Institute, for Applied Economics.

Ominyi, C.N. and Odo, F.A. (2006). Research and statistics in education, management and social sciences. Enugu: Mason Publishers.

1.8 Possible Answers to SAEs

Answers to SAEs 1

1. Research is a systematic investigation to find answers to a research problem or knowledge problem or need.
2. A research proposal is a study plan that is to be followed in the process of research efforts.
3. Researchers are often required to prepare a research proposal for submission to the Research Supervisor or Research Committee, for presentation in a Research Seminar or Workshop, for submission to Research funding agencies and networks to select for funding **(Any 4).**
4. This is a listing of all the major contents of the research proposal.

Answers to SAEs 2

1. A research problem refers to a perceived or felt need which is capable of being resolved or ameliorated by information from a scientific enquiry.

2. The research questions constitute the heartbeat of the research problem.
3. True.

4. The significance section indicates the research outcomes in terms of the contribution to: specific problem, process of policy formulation, and societal development.

5. Descriptive statistics such as frequency, percentage, mean, charts, graphs, partial budgeting, principal component factor analysis and Analysis of Variance (ANOVA) will be employed in data analysis.

Unit 2: RESEARCH REPORT

Unit Structure

2.1 Introduction

2.2 Learning Outcomes

2.3 **RESEARCH REPORT**

2.3.1 Definition of Research Report
2.3.2 Preparation of Research Report

2.4 Target Audience of a Research Report
2.5 Functions of a Good Research Report

2.6 Summary

2.7 Glossary

2.8 References/Further Readings

2.9 Possible Answers to Self-Assessment Exercise(s) within the Content

2.1 Introduction

A report is a written explanation of something that one has observed, heard, done, or investigated. It is a systematic and well organized presentation of facts and findings of an event that has already taken place somewhere or has been found out after an in-depth study has been conducted. Thus, it is only when this is done that we can conveniently say that research output increases the pool of knowledge. This has been discussed fully in this unit.

2.2 Learning Outcomes

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

- Discuss research report;
- Write an essay on how to prepare a research report;
- Identify the target audience of a research report; and
- Evaluate the functions of a good research report.

2.3 Research Report

2.3.1 Definition of Research Report

At the completion of research processes and methodologies, the researcher is expected to put together his research findings in an organized form to enable other scholars and the society to understand and maximize the benefits from the research findings. Report writing is a process of analysing, understanding, appreciating andconsolidating the findings and project a meaningful and coherent view of the phenomenon studied. While a research proposal is a pre-research statement of intent that is ex ante, a research report is a post-research description of work done and the results obtained, that is ex post. A research report is a scientific (objective) assessment for which the researcher should take intellectual responsibility. The sole purpose of scientific writing is to inform, hence it should be accurate and truthful without exaggerations. The report should be factual rather than expressive of opinion, written ina passive rarely containing humour. Since the study is completed when the writing occurs, it is usually written in the past tense although the present tense may be used for those statements of continuing and generalapplicability. The purpose of the report is not communication with oneself but communication with the audience. The author makes an attempt to clarify his own thoughts in order to make it meaningful to the reader and for this purpose, he may have to clearly be aware of what the audience want or need

to know about the study and how the information can be best presented. Reports are used as a form of written assessment to find out what has been learnt from reading, researching or experiencing an important skill that is widely used in the work place. Good report writing is an important quality for any researcher as this also presents the findings to the readers outside your subject area with the experts in the field. It focuses on the findings, conclusions, discoveries made, efforts made and inferences drawn from the research study conducted.

2.3.2 **Preparing Research Report**

A research report will usually be divide into the following three sections: preliminary part, main body of the research and the appendix.

1. Preliminary pages

- **Cover Page:** This is the front page of the research work which contains the title of the project, full names of the author and month and year of the project.
- **Title Page:** This contains the research title, full name of researcher (s), highest qualification at the time of research work, month and year of project.
- **Certification page:** This page confirms that the researcher actually executed the study. It is often signed by the supervisor indicating the month and year the research was completed.
- **Approval Page:** This is a proof that the external examiner, HOD and the research supervisor accepted the project to the department having signed the project report.
- **Acknowledgements:** Here the researcher appreciates those who assisted him on the course of the research.
- **Abstract/Executive Summary:** This contains the summary of objectives, methodology, findings and conclusion drawn. It is usually less than 400 words.
- **Table of Contents:** A table which lists the items in the project.
- **List of Tables (if applicable):** This lists the tables indicating their page numbers.
- **List of Figures (if applicable):** This lists the figures indicating their page numbers.

•**List of appendices:** Presents the appendices as they appear in the main body of the project.

2. Main Body of the Research Report

The main body of a research report consist of the following sections divided into chapters:

•**Introduction: Chapter One.**

a. **Background of the study:** Encapsulates all the relevant historical antecedents to the research problem that will give the audience an insight into the circumstances that have given rise to the problem.

b. **Statement of Research Problem/Research Questions:** Emphasizes the actual problems to which the research will seek explanations.

c. **Objectives of the Research:** Begins with the general statement of intent and ends with the specific objectives.

d. **Statement of Research Hypotheses:** These are tentative answers to research questions.

e. Significance of the study: This indicates the relevance of the research project.

f. Scope/Delimitation of the study: Establishes research limit or boundary.

g. **Operational definition/Definition of terms:** Explains terms and concepts used in the study.

• **Literature Review: Chapter Two**

This consists of four major parts:

a. Conceptual framework

b. Theoretical framework

c. Empirical review

d. Summary of review

The literature review is a vehicle for learning. It is a critical summary assessment of the range of existing materials dealing with the knowledge and understanding on a given subject. It provides an analytical perspective on existing knowledge and provides anchor for the researcher to develop own argument position and analysis. To achieve this goal, it is important to look at previous research on the topic as well as examine the research literature including books, journals and reports.

- **Research Methodology: Chapter Three**

This section consists of the following sections:

a. **Research design:** The purpose of this section is to describe how the research will be carried out.

b. Area of study: States the location of area selected for the research.

c. Population of study: Members of a group under study that are legible for selection.

d. Sample/sampling technique: Segment of the population chosen for a study is the sample while sampling technique is the specific plan adopted for selecting the sample size from a given population.

e. Instrument for data collection: Describes the major features of the questionnaire instrument.

f. Validity of instrument: Explains if the instrument is appropriate for measuring what it is designed to capture.

g. Reliability of instrument: This considers the consistency of the instrument.

h. Method of data collection: Explains the processes used for data collection.

i. Method of data analysis: Reports the statistical tools and techniques used for the study.

- **Results Presentation and Interpretation: Chapter Four**

This summarizes the various tables and charts used for data analysis.

- **Discussion, Conclusion, Implication, Recommendations and Summary: Chapter five**

a. **Discussion:** Here relevant findings are discussed in line with the statistical analysis.

b. **Conclusion:** Contains primarily the major findings of the research.

c. **Implication of the finding:** in a loose sense, this refers to the 'effect' of the study.

d. **Recommendations:** Suggested actions to be taken to bring about improvement based on research findings.

e. **Limitations:** Shortcomings in the research design and execution.

f. **Suggestions for further studies:** Draws attention of other researchers to aspects of the study requiring research attention.

h. Summary of study: Gives a snapshot of the entire research work.

3. **Appendix**

This is the last section of a research report which contains references, questionnaire, statistical results as well as additional tables and figures where applicable.

- **Referencing** is the most vital part of appendices. Referencing is the practice of acknowledging the sources of ideas and information used in a research work, assignment, essay or report.

Scientific knowledge is dynamic and interactive. No scientific work is an island. All scientific works have to be properly and adequately placed in context of existing knowledge. Every research work therefore requires references to the writings and publications of other scientists/authors. It is also necessary to identify these works by making reference to them both in the text and in a list at

the end of the assignment. This practice of acknowledging authors/sources of information is known as referencing. It means that whenever you write an assignment, that requires you to find and use information from other sources, you are expected to refer to these sources. It is the bibliographic listing of information sources at the end of any academic work.

At the end of the research paper, you are required to provide the full bibliographic information (reference) for each source. References must be listed in alphabetical order. Each reference should include four elements; author/editor, date, title and publication information. Information is referenced from books, articles, internet and other print or electronic sources. When you present sentence in a scientific or research paper without a reference or quotation, it means that you are saying to your reader that those ideas, information or words are own original ideas. If they are not, then you may have plagiarized. Plagiarism means copying another's work without acknowledgement or attribution. It is illegal and constitutes a breach of scientific ethics. The purpose of referencing includes:

- Showing the breath of your research;
- Showing the reader the source(s) of your information;
- Strengthening your academic argument;
- Allowing the reader to consult your sources of information; and
- To avoid plagiarism.

Types of Referencing Styles

i. American Psychological Association (APA): This style uses the author-date format to cite references. Example: Booth, W. C., Colomb, G. G. and Williams, J. M. (1995). The Craft of Research. Chicago: University of Chicago Press.

ii. Modern Language association (MLA): MLA style requires all titles and authors' full names not initials to be in italics or underlined. Example: *Berkman, Robert I. Find Fast: How to uncover expert information on any subject. New York: Harper Perennial, 1994.*

iii. Harvard Reference Style: This is an author-date system of referencing. Each reference must be shown each time you use it in the text of the assignment. Example: Daly, L, Mohay, H and Edwards, H 2003, Mother's involvement in Caring for Premature Infant, Elsevier, Sydney.

iv. Footnotes: This is a numerical system of referencing. The footnote system enables you to reference by using consecutive numbers called identifiers within your text starting with the number l. The note identifiers are placed at the end of any quotes, paraphrase, summaries or copied tables, etc. They are written as superscripted Arabic numerals (i.e. numbers slightly raised above the level of the text). Example: F Derham, Art for the child under seven, 7th edn, Australian Early Childhood Association, Watson, ACT 2000.

v. Australian Journal of Physiotherapy (AJP): This is also an author-date style of referencing required in the cases of paraphrase, summary, quote and copy. Example: A recent study (Lim and King 1998) found... or Lim and King (1998) found...

Noteworthy, in writing a research report, whether thesis, dissertation, journal article or conference paper, some fundamental scientific tenets and writing tips should be observed. The research purpose and objectives should be clear, unequivocal and well understood. Concepts should be clarified and placed in proper context. The research design and procedure should be adequate and appropriate for the objectives. Also, the research design and procedures should be reported in such

a manner that allows independent replication of the research, if needed. The research report communicates results in objective, measured and transparent manner. The shortcomings of the research design and the probable consequences on the findings should be discussed. The research findings should be based on evidence and analysis from the research. The researcher should not report or speculate on conclusions not substantiated by the evidence gathered and analysis carried out (Eboh, 2009).

Self-Assessment Exercises 1

1. While a research proposal is a pre-research statement, a research report is…?
2. Suggestions for further studies draws attention of other researchers to…?
3. Outline the three components of the literature review.
4. Mention 4 purposes of referencing includes.
5. State the importance of approval page in a research project.

2.4 Target Audience of a Research Report

The form of a research report will depend upon the target audience. The research report can be prepared for the academic community, the sponsor of a research or the general public. Reports prepared for academic community may take the form of a thesis or dissertation leading to some academic degree or it may take the form of a monograph incorporating in a detailed manner, the entire process of research or it may take the form of a research article dealing with different aspects of the study published independently in different journals. Here, the main objective is to indicate the scope, the methodological variations made in undertaking the study rather than the findings per se.

In the research reports prepared for the sponsors of a research, they become the target audience. The needs of the sponsor may be different as they may want to understand problem or solve a problem. The purpose of the research report is to highlight the various aspects of the problem studied based on the objectives in terms of fulfilling the requirements of the sponsor. It gives answers to the research questions proposed by them in an effort to solve or understand that problem. As for the research reports published for the consumption of the general public, it may take the form of summary reports, articles or brochures. The main concern would be to know the salient features of the findings without bothering about technicalities. Based on the target audience, research reports can be classified into three:

- **Comprehensive research report:** The report should be an effort to narrate the total research process and experience. The experience of the researcher should be documented in selective and organised way. The major thrust of the report should be to communicate what actually happened throughout the research process and not simply what we hoped would happen at each

stage. The report should not be merely selective – narration of our successes but may also contain the limitations along with the reasons for it.

- ***Research articles:*** *It deals with different facets of a research problem in a more detailed fashion, emphasising the methodology adopted in the study. Also, the conclusions are to be related to the specific objectives.*

- ***Summary report:*** *These are generally meant for the general public. Since summary reports may receive wide distribution, they are frequently written in less technical language and may make greater use of better methods ofpresenting the data. It focuses on major findings, their elucidation in simple and clear language and categorical statements about theimplication of the study and recommendations.*

Each of these may take the form in which it meets the requirements of the target audience. However, it is assumed that attempting to write a multi-purpose report to serve all audience becomes difficult because what each one looks for in the report will be different. It is assumed that the most useful approach is to choose and write for a single target audience at a level appropriate to it. The form, content and style of the research report should be chosen to suit the level of knowledge, experience and interest of the target audience as well as anticipate the likely uses for the research findings.

2.5 Importance of a Good Research Report

Research work is an exciting thing to execute and it has numerous benefits both in the immediate and for the future of students and society at large. For the undergraduate and postgraduate students, a good research report earns good grades for students as it accounts for about 40% of the required academic work load for most students. Students who receive poor grades in research projects are often at risk of obtaining poor final grades on graduation. Those who obtain outright fail in research

projects spend extra academic sessions so as to remedy it. However, the importance of research report goes beyond obtaining grades. The knowledge and experience gained while painstakingly working on a research report will go a long way in helping the student in his future academic and social endeavours. An outstanding research report is has the possibility of winning an award for the researcher. Research reports when published in journals and brought to public domain serve as sources of material for future publications and by extension major contributions to knowledge. Unarguably, individuals and organisations expect to benefit much from research findings: discoveries, concisions, suggestions and solutions to problems, new ideas, and recommendations made by researchers in their research reports. Research is a good and rewarding exercise for students and helps them to achieve immediate goals as well as prepare them for a bright future. Published research reports have a tendency of out-living their authors if it is well-preserved. Basically, research reports have the following importance among others:

- To provide the information regarding the findings of research work i.e. methods, data analysis, conclusion and so on in the systematic, scientific and accepted way.
- To elicit crucial facts for solution derived and decision making.
- To prove the worth and legitimacy of assigned research job.
- To provide the judgment tools for the judgment of quality and talent of researcher within and outside the academia.
- To communicate the research findings professionally.
- To pertain the credibility of the research.
- To develop appreciation of standards, consolidate arguments and identify the knowledge gaps.

Self-Assessment Exercises 2

1. Plagiarism means…..?
2. Based on the target audience, research reports can be classified into three….?
3. References must be listed in …… order.
4. List the sections in the chapter five of a research project.
5. State the purpose of research design.

2.6 Summary

The method of organising the research findings and experience in a coherent manner constitutes the research report. While writing the report, the researcher should keep in mind the target audience: academic community, administrators and policy makers or the general public. In this unit, we have learnt how to write the report of a research. We have been able to look at the different audience and their needs in terms of the structure and detail expected in a research report. Suggestions were also offered for preparing comprehensive research report as can be seen in this unit.

2.7 References/Further Readings

Eboh, E. C. (2009). Social and Economic Research: Principles and Methods. African Institute for Applied Economics, Enugu.

Fernandes, M. (2009). Statistics for Business and Economics. Marcelo Fernandes and Ventus Publishing APS, Aps. www.bookboon.com. Accessed 4th September, 2022.

Kara, O. (2012). Researching for Hospitality and Tourism Management. An unpublished course material of the Uttarakhand Open University.

Ominyi, C.N. and Odo, F.A. (2006). Research and statistics in education, management and social sciences. Enugu: Mason Publishers.

2.8 Possible Answers to SAEs

Answers to SAEs 1

1. While a research proposal is a pre-research statement, a research report is a post-research description of work done and the results obtained.

2. Suggestions for further studies draws attention of other researchers to aspects of the study requiring research attention.

3. This consists of three major parts:

   a. Conceptual framework

   b.Theoretical framework

   c.Summary of review

4. The purpose of referencing includes:

- Showing the breath of your research;
- Showing the reader the source(s) of your information;

- Strengthening your academic argument;
- Allowing the reader to consult your sources of information; and
- To avoid plagiarism **(Any 4).**

5. Approval Page: This is a proof that the external examiner, HOD and the research supervisor accepted the project to the department having signed the project report.

Answers to SAEs 2

1. Plagiarism means copying another's work without acknowledgement or attribution.

2. Comprehensive research report, Summary report and Research articles.

3. References must be listed in alphabetical order.

4. Sections in the chapter five of a research project are: Discussion, Conclusion, Implication, Recommendations and Summary.

5. Research design: The purpose of this section is to describe how the research will be carried out.

UNIT 3 Application of Statistics and Computer in Research.

Unit Structure

3.1 Introduction

This unit provides you with a general understanding of the relevance of the application of statistics and computer in research. We also identified the different computer programmes for different types of data analysis. It is hoped that at the end of the unit you would have achieved the objectives stated herein.

3.2 Learning Outcomes

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

- Discuss the roles of statistics in research; and

- Evaluate the computer software applied to research.

3.3 Application of Statistics and Computer in Research

3.3.1 Role of Statistics in Research

The application of statistics to research plays the following roles:

i. Easy comprehension and interpretation of large data: Statistics simplifies complex mass of data and presents them in a comprehensive way that they are at once made easy to comprehend and interpret. Instead of having a large raw data, the data are prepared in percentages and means which can be grasped more easily than a mass of data. Also, statistical analyses in the form of histogram, bar chart, or pie chart, make it easier for you to understand.

ıı. **Easy conclusion**: Statistics presents data in more comprehensive and definite form: Statistics made conclusion to be stated numerically to be more convincing than conclusions stated qualitatively. For example, it is more attractive and convincing to say 85% of candidates that sat for WAEC passed in 2021 than saying most candidates that sat for WAEC in 2021 passed.

ııı. A**ttractiveness in presentation of figures and facts:** Statistics interpret conditions that are more presentable. Statistics present conditions in an attractive ways such as pie chart and histogram or bar charts of the phenomenon under investigation. Also, certain conditions are proved statistically to find out the probability of future occurrence of such situation so that necessary actions could be taken to prevent future occurrence of such conditions.

iv. **Easy classification of numerical data:** It provides easy way of classifying numerical data. The method of classification in statistics provides the salient features of the variables that is under consideration. For example, statistical methods provide appropriate method of classifying two or more data by bringing out the maximum, minimum and the standard deviation of the various categories.

v. **Easy Comparison of data**: It provides an easy way of comparing data. Some data may be meaningless unless they are subjected to statistical analysis before they can be compared with similar data at other places. Statistics made an easy way of relating two different masses of numerical data by comparing some relevant information from the two sets of data such as comparing their means, medians and modes of their distribution.

3.3.2 Application of computer in research

Data analysis and interpretation of research results have been made easy world over due to the sophistication and advancement in science and technology. The application of information and communication technology have impacted positively on the research process and data analysis in particular. Many computer programmes for data analysis and interpretation are now at the disposal of researchers. For instance, Statistical Package for Social Scientists (SPSS) is a computer software for data analysis in social and behavioral sciences. This software accepts raw data, analyses them and produce results as required. Table 27 below shows different computer programmes for different types of data analysis.

Table 27: computer programmes for data analysis

S/N	Statistical procedures	Computer software
1.	Univariate Analysis (Distribution, central tendency and dispersion)	Microsoft Excel, SPSS, SAS, STATA etc.

2.	Bivariate Analysis (Correlation and regression analyses)	Microsoft Excel, SPSS, SAS, STATA, Economic Views, PC Give, PC Fml. Give Win. etc.
3.	Inferential Statistics (T-Test, ANOVA, ANCOVA etc.)	Microsoft Excel, SPSS, SAS, STATA, LIMPED, etc.
4.	Inferential Statistics (Chi-square, Normal Distribution, etc.)	SPSS, SAS, STATA, LIMPED, etc.
5.	Models of Qualitative Choice (Probit, Logit, Tobit)	SPSS, SAS, STATA, E-Views.
6.	Multivariate Analysis (Principal Component Analysis, Cluster Analysis and Discriminant Analysis)	SAS, STATA, LIMPED, etc.

Self-Assessment Exercises 1

1. Write the full meaning of SPSS.
2. List the computer software for analyzing ANCOVA.
3. SPSS, SAS, STATA and E-Views are used to analyse ….?
4. Each reference should include four elements…..?

3.4 Summary

It is obvious that no reliable scientific research results can be obtained without the application of statistics and computer software. SPSS and other computer software is highly recommended for use in any serious research undertaking no matter the cost. However, researchers are advised to obtain quality data which must be properly coded before feeding into the computer for high precision.

3.5 Glossary

APA: American Psychological Association

AJP: Australian Journal of Physiotherapy

SPSS: Statistical Package for Social Scientists

MLA: Modern Language Association

3.6 References/Further Readings

Eboh, E. C. (2009). Social and Economic Research: Principles and Methods. African Institute for Applied Economics, Enugu.

Fernandes, M. (2009). Statistics for Business and Economics. Marcelo Fernandes and Ventus Publishing APS, Aps. www.bookboon.com. Accessed 4th September, 2022.

Kara, O. (2012). Researching for Hospitality and Tourism Management. An unpublished course material of the Uttarakhand Open University.

Ominyi, C.N. and Odo, F.A. (2006). Research and statistics in education, management and social sciences. Enugu: Mason Publishers.

3.7 Possible Answers to SAEs

Answers to SAEs 1

1. SPSS: Statistical Package for Social Scientists.

2. Microsoft Excel, SPSS, SAS, STATA, LIMPED etc.

3. Models of Qualitative Choice (Probit, Logit, Tobit).

4. Each reference should include four elements: author/editor, date, title and publication information.