NATIONAL OPEN UNIVERSITY OF NIGERIA

SCHOOL OF MANAGEMENT SCIENCES

COURSE CODE: BHM 844

COURSE TITLE: RESEARCH METHODS FOR BUSINESS DECISIONS

COURSE GUIDE

BHM 844
RESEARCH METHODS FOR BUSINESS DECISIONS

Course Team	Dr. Onyemaechi J. Onwe (Course Writer/Developer) -NOUN Dr. Onyemaechi J. Onwe (Programme Leader) - NOUN Mrs. Inua (Course Coordinator) - NOUN

NATIONAL OPEN UNIVERSITY OF NIGERIA

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

Abuja office
NOUN Building
5 Dar es Salaam Street
Off Aminu Kano Crescent
Wuse II, Abuja

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

Published by
National Open University of Nigeria

Printed 2013

ISBN: 978-058-868-X

CONTENTS **PAGE**

INTRODUCTION

The course, BHM 844: Research Methods for Business Decisions is a two-credit course for post-graduate diploma (PGD) students in the School of Business and Human Resources Management.

The course has been systematically arranged for you in six modules comprising of 15 distinct but related units of study activities. This course guide informs you of what you need to know about the aims and objectives of the course, components of the course material, arrangement of the study units, assignments, and examinations.

THE COURSE AIM

This course is aimed at acquainting students with the practical aspects of research in humanities and social sciences. Its major focus is on the globally acceptable research process, including data collection, data analysis, and statistical inferences. To ensure the achievement of this aim, some important background information have been provided and discussed. These are:

- definition of research and the research process
- essential parts of a research project
- the research design
- the research proposal
- research questions and hypotheses
- data and sources of data
- methods of data collection, presentation, and analysis
- correlation analysis
- regression analysis
- determination of statistical sample size.

COURSE OBJECTIVES

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

- define and understand the research process
- choose a researchable project topic
- write a good research proposal
- enumerate the different types of research designs
- define and list the types of sampling designs
- determine the sample size for a given survey activity
- write research questions and construct the corresponding questionnaires

- list the characteristics of a good questionnaire, and administer such questionnaire
- effectively present data, analyse them, and prepare the research report.

COURSE MATERIALS

The course material package is composed of:

1. Course Guide
2. Study Units
3. Self-Assessment Exercise
4. Tutor-Marked Assignment
5. References/Further Reading

STUDY UNITS

The study units of the course grouped under different modules are shown below:

Module 1 Organisation of the Research Process

Unit 1 Introduction to Research Process
Unit 2 Essential Parts of Research Projects

Module 2 The Research Design

Unit 1 The Research Design
Unit 2 Components of a Research Design

Module 3 Research Proposals

Unit 1 The Research Proposal
Unit 2 Structuring and Evaluating a Research Proposal

Module 4 The Research Hypotheses

Unit 1 Research Questions and Hypotheses
Unit 2 Hypotheses Testing I
Unit 3 Hypotheses Testing II

Module 5 Data Presentation and Analysis

Unit 1 Data, Sources of Data, and Methods of Data Collection
Unit 2 Data Presentation and Analysis

Module 6	**Tools of Data Analysis and Determination of Sample Size**
Unit 1	Advanced Parametric Tools and Data Analysis: The Linear and Rank Correlation Coefficients
Unit 2	Advance Parametric Tools of Analysis: The Partial Correlation Coefficients
Unit 3	Advanced Parametric Tools of Analysis: The Simple Linear Regressions
Unit 4	Population, Sampling, and Determination of Sample Size

SELF-ASSESSMENT EXERCISE

Each unit of the course has a self-assessment exercise. You will be expected to attempt them as this exercise will enable you understand the content of the unit.

TUTOR-MARKED ASSIGNMENT

The tutor-marked assignment (TMA) at the end of each unit is designed to test your understanding and application of the concepts learned. It is extremely important that you submit these assignments to your facilitators for assessments and comments. Tutor-marked assignment scores consist of 30% of the total grading score for the course.

FINAL EXAMINATION AND GRADING

At the end of the course, you will be expected to participate in the final examinations as scheduled. The final examination constitutes 70% of the total grading score for the course.

SUMMARY

The course, BHM 844: Research Methods for Business Decisions is ideal for today's business decisions as well as business research activities. It will enable you carry out research activities in such managerial functions as planning, controlling, forecasting, evaluations, and reporting. Having successfully completed the activities as required by the course, you will be equipped with the global expectations of business research activities and related business decisions. You will find the course very useful in your daily business, economic, and social decisions.

MAIN COURSE

CONTENTS **PAGE**

MODULE 1 ORGANISATION OF THE RESEARCH PROCESS

UNIT 1 INTRODUCTION TO RESEARCH PROCESS

CONTENTS

1.0 INTRODUCTION

Unit 1 introduces you to the basic concepts of research. It presents important definitions which are required for the basic understanding of a scientific research process.

2.0 OBJECTIVES

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

- give the definitions of research
- explain a research process
- analyse some basic research concepts.

3.0 MAIN CONTENT

3.1 Definitions of Research

Research can simply be defined as the process of arriving at dependable solutions to problems through planned and systematic collection, analysis and interpretation of data.

Research is an important tool for advancing knowledge for promoting progress, and for enabling scholars to relate more effectively to the environment, to accomplish their objectives, and to resolve conflicts.

Research is often based on two types of definitions: conceptual definition and operational definition. **Conceptual definitions** are definitions that describe concepts by using other concepts. For example, a conceptual definition for "political violence" might be an "aggressive behaviour toward political institutions and persons occupying political posts." One conceptual definition of "intelligence" might be "the ability to think in an abstract manner." Another might be "the ability to solve problems."

Operational definitions attempt to bridge the gap between the theoretical-conceptual level and the empirical-observational level. An operational definition involves a series of instructions describing the operations that must be carried out by a researcher in order to demonstrate the existence, or the degree of existence, of an empirical occurrence represented by a concept.

Many concepts in the social sciences and humanities are given operational definitions solely on the strength of reactions to specific situations, since manipulation of the property to be defined is often difficult. A research will often argue that a certain individual is "conservative" if he or she answers a series of questions in a specific manner. The assumption is that certain answers to specific questions represent particular personality patterns, one of which is "conservatism."

Research attempts to discover relationships existing among important variables. It is aimed at finding the condition under which a certain phenomenon of interest occurs and the conditions under which it does not occur in what might appear to be similar circumstances.

3.2 Basic Research Concepts

We look at the two often used research concepts:

(i) the variables

(ii) the hypotheses

3.2.1 The Variables

A variable is an empirically applicable concept that takes on two or more values. Examples of empirically applicable concepts that are treated as variables include: "social class"; "expectations"; "tolerance;" "political participation;" and, "membership in organisations." For instance, social status may be symbolised by the letter "Y" and differentiated by at least five values such as: lower, lower middle, middle, upper middle, upper. Income may be symbolised by the Letter "I" and graded with three values: low, medium, high. Income can also be graded by a series of numerical, for example, N100,000, N250,000, and N500,000.

Observe that some variables have only two values. For example, if the empirically applicable concept is "sex", then its values can only take female and male. Other examples of two-value variables are: Middle Class-Working Class, Student-Non student. These types of variables are often referred to as being dichotomous variables. Most variables investigated in the social sciences are however, not dichotomous. They are instead, characterised by a large number of values.

There are three common types of variables in scientific research, whether dichotomous or multi-valued. These are:

(i) Independent variables or predictor variables;
(ii) Dependent variables or criterion variables; and,
(iii) Control variables or test variables.

The variable that the researcher is interested in explaining is the dependent variable. The explanatory variables are the independent variables. An independent variable is the hypothesised cause of a dependent variable, and the dependent variable is the expected outcome of the independent variable.

The distinction between the above types of variables is analytic and relates to a particular purpose of research. An independent variable in one study may be a dependent variable in another. The decision to treat a variable in terms of the above types of variables will depend on the objective of a given research. Nevertheless, after making a decision, the researcher has to be consistent in his or her classification throughout the research process.

You can illustrate the relationship between an independent variable and a dependent variable using a two-dimensional graph. Following the mathematical custom, X can be referred to as the independent variable and you can represent it by the horizontal axis, while Y, the dependent variable, can be represented by the vertical axis. The X-values are plotted on the X-axis, and Y-values on the Y-axis. For example, suppose that in a study of revenues obtained by a profitable business venture, you have two sets of measures: X which measures the number of units sold per sales period, and Y which measures the corresponding revenues from sales. Table 1 shows some hypothetical data on the two measures, X and Y.

Table 1: Number of Units Sold of a Product and Revenue Obtained from Sales

Number of Units Sold (X)	**Revenue Obtained (N'000s) (Y)**
1	5
2	8
3	12
4	14
5	14
6	18
7	22
8	20
9	25
10	28

You can refer to table 1 as the revenue schedule. You can attempt plotting this schedule on the two-dimensional graph.

3.2.2 The Hypotheses

Hypotheses are regarded as tentative answers to research problems. They are usually expressed in the form of a relationship between independent and dependent variables. Hypotheses are said to be tentative because their validity are evaluated only after they have been empirically tested. When you as a researcher propose a hypothesis, you will lack assurance that it will be verified. Researchers construct a hypothesis, and if it is rejected, they consider another hypothesis.

You can come up with a hypothesis either by deduction from theories, or directly from observations, or by intuition, or from a combination of these. The source of a hypothesis is of little significance compared with the way in which it can be rejected or accepted.

If hypotheses are rejected, there is a need to modify the theory from which they were deduced. The failure to reject hypotheses does increase the credibility of the theory. Where formal theories are not available for hypothesis deduction, hypothesis can be generated from conceptual frameworks. In such cases, the failure to reject a hypothesis may lead to the construction of a more systematic and rigorous theory.

3.2.3 Characteristics of a Researchable Hypothesis

For a hypothesis to be researchable, no matter its source, it must possess the following characteristics:

1. The hypothesis must be clear: Clarity can be obtained by means of definitions. You require valid operational definitions for all the concepts in the research hypothesis. In the defining process, you should use professional literature and experts' opinions.
2. Given that research in the social sciences is to some extent a social activity whose problems and methods tend to be affected by the milieu in which it takes place, you as the researcher must be aware of your values and make them as explicit as possible
3. The hypothesis must be specific. You must explicate the expected relations between the variables and the conditions under which these relations will hold. For instance, a hypothesis stating that "X is related to Y" is over-generalised and will not allow concrete predictions. This is so because the relationship between X and Y can be positive or negative.
4. For hypotheses to be researchable, you must make sure that there are methods available for testing them.

The scientific approach to research relies on the observations and methods employed in generating observations. This approach assumes that the world can be known only as experience is processed through human intelligence.

3.3 Basic Operation in Scientific Research Process

Scientific explanations in the research process, whether deductive or probabilistic, explicate the factors in a situation that are responsible for the occurrence of a particular phenomenon. In practice, this involves four distinct operations:

1. **Demonstrating co-variation*:*** Co-variation simply means that two or more phenomena (or variables) vary together. For example, if a change in the level of education is accompanied by a change in productivity (or income), one can say that education co-varies with productivity (income). In scientific research, the notions of co-variation are expressed through measures of relationships, commonly referred to as correlations or association.

Thus, a correlation between variables is necessary evidence for a casual interpretation.

2. **Eliminating spurious relations:** This operation requires the scientist to demonstrate that the observed correlation is non-spurious. A non-spurious relation is defined as a correlation between two phenomena (variables) that cannot be explained by a third factor. In other words, if the effects of all relevant factors are eliminated and the relation between the investigated phenomena is maintained, then the relation is non-spurious.
3. **Establishing the time order of occurrences:** This requires the researcher to demonstrate that one phenomenon occurs first or changes prior to another phenomenon. For example, studies have shown that the correlation between urbanisation and democratic development is non spurious. To establish that urbanisation is casually related to democratic development, it must also be demonstrated that the former precedes the later.
4. **Theorising:** Theory is viewed here as an interpretation of, or conceptual justification for, an observed co-variation. This interpretation specifies the casual nature of a co-variation by explicating the mechanism that connects the phenomenon under investigation.

3.4 The Research Process

The research process is the overall scheme of scientific activities in which research scientists engage. Seven principal types of activities are enumerated below, each of which can be regarded as a process in the research process.

i. **Problem identification:** A problem is an intellectual stimulus calling for a response in the form of scientific solution. The first stage in any research process is to identify the research problem(s).

ii. **Statement of hypotheses about the identified problem(s):** Stimuli (or problems) that are too general or too abstract present difficulties in the investigation and therefore requires concretisation. This is attained by hypotheses. Hypotheses are regarded as tentative answers to researchable problems. The researcher breaks down a general problem into a set of concrete hypotheses and investigates each separately. Consider the problem, “What brings youth to universities?” This involves considerations such as the motivation to study, the individual’s financial state, the social background of one’s family, encouragement by peers, and one’s academic achievements in secondary school. These five considerations generate five hypotheses:

- the inclination to enroll at university increases with the motivation to study
- an individual's financial state co-varies with his or her propensity to study at the university
- youth from relatively well - to-do families tend to enroll at universities
- the greater the encouragement to study at university given by peers, the stronger the tendency to do so
- academic achievement in secondary school co-varies with the tendency to study at university.

Each of these hypotheses can be tested by observation on the basis of which it will either be rejected or accepted. The rejection option is the characteristic feature of hypotheses and the reason for viewing them as tentative answers.

iii. **Research design:** A research design is the structure, programme, and strategy upon which hypotheses are evaluated.

iv. **Measurement:** Measurement can be defined as any procedure whereby observations are systematically assigned symbols. These symbols are amenable to logical, mathematical and statistical manipulations that reveal information that otherwise could not have been revealed. Symbols can be added, subtracted, percentage, introduced as subjects or objects in sentences, or employed as elements in graphs and diagrams.

v. **Data collection:** This is the stage whereby observations are made and recorded. Data can be collected in several ways including field observations and survey (for primary data), and journals and publications (for secondary data). Any data collection method can also be used to gather information concerning different hypotheses.

vi. **Data analysis:** At this stage, all available data is analysed according to research objectives, research questions, and hypotheses. The analytical method to be used will depend basically on the type of research as well as the research design.

vii. **Empirical generalisations:** An empirical generalisation is a statement asserting a universal connection between variables of interest. The logic whereby observations are transformed into empirical generalisation is referred to as **induction**. We may distinguish two types of inductive principles: one is the principle of enumeration in which inductive hypothesis is treated as being well established if it has not been refuted by experience. The other inductive principle is that of elimination, according to which an inductive hypothesis is taking to be well established if, while it has not been refuted by experience, alternative hypothesis have been so refuted.

SELF-ASSESSMENT EXERCISE

In a research process, discuss how the term 'constant' differs from a variable.

4.0 CONCLUSION

You have been informed on the theoretical definition of scientific research. You must have also learned about what variables are and what research hypotheses stand for.

5.0 SUMMARY

Research is based on two types of definitions: conceptual definition and operational definition. Conceptual definitions are definitions that describe concepts by using other concepts. Operational definitions attempt to bridge the gap between the theoretical-conceptual level and the empirical-observational level. An operational definition involves a series of instructions describing the operations that must be carried out by a researcher in order to demonstrate the existence, or the degree of existence, of an empirical occurrence represented by a concept.

A variable is an empirically applicable concept that takes on two or more values. There are three common types of variables in scientific research, whether dichotomous or multi-valued. These are:

(i) Independent variables or predictor variables
(ii) Dependent variables or criterion variables
(iii) Control variables or test variables.

Hypotheses are regarded as tentative answers to research problems. They are usually expressed in the form of a relationship between independent and dependent variables. Hypotheses are said to be tentative because their validity are evaluated only after they have been empirically tested.

The research process is the overall scheme of scientific activities in which research scientists engage. Seven principal types of activities in the research process include: (i) problem identification; (ii) statement of hypothesis about the identified problem; (iii) research design; (iv) measurement; (v) data collection; (vi) data analysis; and, (vii) empirical generalisation.

6.0 TUTOR-MARKED ASSIGNMENT

Enumerate the activities necessary in a scientific research process, and explain why they are important.

7.0 REFERENCES/FURTHER READING

Nachimias, D. & Nachimias, C. (1976). *Research Methods in the Social Sciences.* New York: Martin's Press.

Nworgu, B. G. (1991). *Educational Research: Basic Issues and Methodology.* Ibadan: Wisdom Publishers Ltd.

Osuala, E. C. (1993). *Introduction to Research Methodology*. Onitsha: Africana-Fep Publishers Ltd.

UNIT 2 ESSENTIAL PARTS OF A RESEARCH PROJECT

CONTENTS

1.0 INTRODUCTION

Students often have difficulties in the arrangement of the write ups of a research project. In recognition of these difficulties, this unit puts together the necessary parts of a research project. The unit should guide you in presenting your research work.

2.0 OBJECTIVES

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

- choose a researchable topic
- write a research proposal
- define and understand the various types of research hypotheses.

3.0 MAIN CONTENT

3.1 Choosing a Research Topic

To most students, the most difficult aspect of a research project or dissertation is the choice of a research topic. Some students abandon a chosen topic after many hours of exploration, while others continue even though the research problem is unsuitable and end up having nothing worthwhile but the satisfaction of having met another requirement for the award of a degree. It is important you begin by identifying the sources of a research topic and guidelines for choosing a research topic.

3.1.1 Sources of Research Topics

The research student may be guided by the following sources of research topics:

(i) **Observations and experience:** A research topic can evolve from personal experiences and observation of a set of problems. Such incidents as fire outbreak, theft, loss of sales, business distress and the like, are prospects for research topics.

(ii) **Review of literature:** Literature is one of the major sources of research topics. A review of literature on subjects of interest can reveal several important issues and problems that call for research. Topical discussions in classroom subjects can also present excellent opportunities in the choice of a research topic.

(iii) **Previous research projects:** As part of their recommendations, many past research projects often make recommendations for future research. These recommendations are also excellent sources of research topics.

(iv) **Theories:** These are propositions explaining certain phenomena, such as the theory of demand and supply. The propositions here are good sources of research topics.

3.1.2 Guidelines for Selecting a Research Topic

Although there are no standard rules that will ensure the suitability of a research problem, a number of suggestions can be made as guides in the choice of a research topic.

(i) The topic must be of personal interest to the researcher

(ii) The topic should be sufficiently original that it does not involve objectionable duplication

(iii) The topic must be researchable

(iv) The topic must be significant i.e. must be capable of contributing to existing knowledge

(v) The research into the problem must be feasible i.e. must ensure that data are available.
(vi) The topic must be consistent with the researcher's competence, interest and circumstance.

3.2 Suggested Useful Questions in the Choice of a Research Topic

The followings are useful questions that will aid in the choice of a research topic:

(i) In your field of interest, what practical problems do you think have to be met by those individuals doing the actual work?
(ii) What problems are under active attack in the recent research?
(iii) What facts, principles, generalisations and other findings have resulted from research in your field?
(iv) What practical implication for school, work can you draw from the results?
(v) To what extent have the findings of research actually been applied to your field?
(vi) What problems remain to be subjected to research and what problems are now emerging?
(vii) What are the difficulties to be met in prosecuting the research yet to be conducted in your field?
(viii) What are the relationships between research in your field and research in adjacent field?
(ix) What research techniques or procedures have been developed in your field?
(x) What concepts are being operative, either explicitly or implicitly in the research in your field?
(xi) What assumptions have been implicit in the research in you field?

3.3 Research Proposal

A research proposal is analogous to the plans and specifications, which precedes the construction of a building. The whole must be envisaged and each detail must be conceived.

A research proposal should contain the following parts:

(a) **Title:** The title of the research (or research topic) comes first in the proposal. The choice of terms in the title should help to indicate the scope of the research.
(b) **Introduction:** This serves as the backbone of the research. Here, the researcher explains how he or she became interested in the problem and how he or she felt the study is important, establishes the need for the study and identifies its purpose.

(c) **Statement of the problem:** The statement of the research problem should elaborate upon the information implied in the title of the research. The problem statement should be brief and in specific terms.

Research problem can take one of the following forms:

(i) An unsatisfactory state of affairs – state that arises due to deviations from expectations.
(ii) An unanswered question – those questions demanding some answers constitute research problems.
(iii) A missing link – a gap in the implementation process.
(iv) An un-met need – existence of need yet to be satisfied constitutes a research problem.
(v) Organisational imbalance – imbalance in the development and provision of facilities in an organisation can constitute a problem.

The points to be borne in mind in stating a research problem are that the researcher should:

i. Know what a problem is
ii. Tell the reader what the research problem is all about
iii. Learn and adopt certain characteristics of problem statements.

(d) **Purpose of research:** The purpose of the research should be a quick overview of the research itself. The purpose should be written in clear and concise manner, indicating the important aspects of the research.

(e) **Hypotheses:** This may or may not be included in the proposal, depending on the type and nature of the research. Hypotheses are usually incorporated if statistical testing is to be used. If needed, the hypothesis should be stated as null hypothesis – a statement that no significant relationship between the variables exists.

Notice that a hypothesis is an informed, intelligent guess about the solution to a problem. It is a proposition whose validity needs to be established.

3.3.1 Importance of Hypotheses

1. Provides a guide and direction to the research.
2. Draws the attention of the researcher to the important aspects of the problem under investigation.
3. Provides a framework for drawing conclusions.

3.3.2 Types of Hypotheses

There are basically two types of hypotheses in social science research

(i) **Research hypotheses:** These are often referred to as non–parametric hypotheses. They are postulation about the relationships between two or more variables that are of critical interest in the solutions of the research problem. Research hypothesis does not express the variables in measurable terms.

(ii) **Statistical hypotheses:** These hypotheses, often referred to as parametric hypotheses, are propositions about statistical population which is to be verified on the basis of data collection from a sample of the given population. These hypotheses express the relationship between two or more variables in statistical and measurable terms.

3.3.3 Sources of Hypotheses

Sources of hypotheses include:

i. Experience
ii. Literature
iii. Theory
iv. Previous Findings

3.3.4 Characteristics of Good Hypotheses

A good hypothesis should be:

i. Testable
ii. A statement of an expected relationship between two or more variables
iii. Plausible
iv. Consistent with current knowledge
v. Unambiguous.

3.4 Other Parts of a Research Proposal

Significance of the research: This should contain the importance and usefulness of the research.

Scope and limitations of the research: The scope of the research sets forth the exact bound of the topic being researched. Scopes are restrictions imposed on the research by the researcher himself. Limitations are restrictions imposed on the research by the nature of the research itself. These limitations serve as constraints to the research and can affect the result of the study.

Definition of terms: Terms that have unique use in the study should be defined in the research proposal.

Review of related literature: The review of literature serves two purposes in the research report.

(i) It set the theoretical base for the research
(ii) It set the current research into perspective to show "the state of the art."

The review of related literature should be in an organised form, with appropriate sub–heads to indicate the areas or topics covered. Authors must be appropriately cited with the correct referencing format. (For acceptable referencing formats, see Onwe, O.J. (1998).

Research design: You need to specify your research design when writing a research proposal. This is a plan or blueprint specifying how data relating to the research will be collected and analysed. Research design helps in the following ways:

i. It provides the procedural outline for the research.
ii. Provides the researcher with the necessary framework for tackling a particular research problem.
iii. Helps in the proper delineation of the scope and limitations of the research.
iv. Acquaints the researcher with the potential problems in the execution of the research.

3.5 Presentation and Analyses of Data

This is where the research data is presented and analysed for possible inferences. This is a chapter of its own and has three important sections:

(i) Presentation and analysis of data according responses to the research questions.
(ii) Presentation and analysis of data on test of hypotheses.
(iii) Presentation and analysis of other relevant data.

Data must be presented objectively, awaiting editorial comments or conclusions about what the data shows. The analytical tables should be presented with appropriate numbering, titles and sources.

3.6 Summary of Findings, Recommendations and Suggestion for Future Research

The summary serves as a synopsis which many readers will read first in order to determine if the research is worth further readings. The procedures should be summarised in general terms with only enough detail given for the reader to obtain a general picture of what was done.

Recommendations are made based on the research findings. Most research studies seem to raise more questions than they answer and these questions lead to recommendations for future research.

3.7 Abstracts

The abstract is a brief summary of the research, usually about 100 words in length or one page. It should give information on the research problem that was studied, the research methodology, the results and the major conclusions.

A good abstract should have approximately four paragraphs made up of:

(i) Two or three statements or sentences about the problems and objectives of the research.
(ii) Two or three statements or sentences about the research method that was used.
(iii) Two or three statements about the research findings.
(iv) Two or three statements or sentences about the recommendations and conclusions based on the research finding.

SELF-ASSESSMENT EXERCISE

Using all you have learned from this unit, choose and introduce two researchable topics.

4.0 CONCLUSION

This unit has exposed you to the major areas or parts of a research project, including: the project title, the introductory part, literature review, and research methodology. The unit also discussed in detail the basic requirements in the choice of a research topic, including understanding of the sources of research topics.

5.0 SUMMARY

To choose a research topic, you have been advised to begin by identifying the sources of a research topic and guidelines for choosing a research topic. The useful sources include: observation; literature review; previous research; and theory.

Among the guidelines in choosing a research topic are: the topic must be of personal interest to the researcher; the topic should be sufficiently original that it does not involve objectionable duplication; the topic must be researchable; the topic must be significant; the topic must be capable of contributing to existing knowledge; the research into the problem must be feasible; the research should ensure that data are available; and, the topic must be consistent with the researcher's competence, interest and circumstance.

Other major issues discussed in the unit are the importance and types of hypotheses. Regarding the types of hypotheses, you learned two of them including (i) research hypotheses; and, (ii) statistical hypotheses.

6.0 TUTOR-MARKED ASSIGNMENT

Explain why it is important to understand the major parts of a research proposal.

7.0 REFERENCES/FURTHER READING

Cooper, D. R. & Schindler, P. S. (2001). *Business Research Methods*. New York: McGraw-Hill.

Onwe, O. J. (1998). *Elements of Project and Dissertation Writing: A Guide to Effective Dissertation Report.* Lagos: Impressed Publishers.

Nworgu, B.G. (1991). *Educational Research: Basic Issues and Methodology*. Ibadan: Wisdom Publishers Ltd.

Osuala, E. C. (1993). *Introduction to Research Methodology*. Onitsha: Africana-Fep Publishers Ltd.

MODULE 2 THE RESEARCH DESIGN

UNIT 1 THE RESEARCH DESIGN

CONTENTS

1.0 INTRODUCTION

The previous unit presented a brief discussion on the concept of research design. In that unit, the discussion mentioned research design as an important part of the research process. In this unit, a much more detailed discussion on research design will be presented to you. Research design can be thought of as the structure of research -- it is the "glue" that holds all of the elements in a research project together. In scientific research process, we often describe a design using a concise notation that enables us to summarise a complex design structure efficiently. What are the "elements" of a good design? They are:

1. Observations or measures

These are symbolised by an 'O' in design notation. An ‘O’ can refer to a single measure (e.g., a measure of body weight), a single instrument with multiple items (e.g., a 10-item self-esteem scale), a complex multi-part instrument (e.g., a survey), or a whole battery of tests or measures given out on one occasion. If you need to distinguish among specific measures, you can use subscripts with the O, as in O_1, O_2, and so on.

2. Treatments or programmes

These are symbolised with an 'X' in design notations. The X can refer to a simple intervention (e.g., a one-time surgical technique) or to a complex hodgepodge programme (e.g., an employment training programme). Usually, a no-treatment control or comparison group has no symbol for the treatment (some researchers use X+ and X- to indicate the treatment and control respectively). As with observations, you can use subscripts to distinguish different programmes or programme variations.

3. Groups

Each group in a design is given its own line in the design structure. If the design notation has three lines, there are three groups in the design.

4. Assignment to group

Assignment to group is designated by a letter at the beginning of each line (i.e., group) that describes how the group was assigned. The major types of assignment are:

R = random assignment
N = nonequivalent groups
C = assignment by cutoff

2.0 OBJECTIVES

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

- give various classifications of research design
- familiarise yourself with the different types of research design
- highlight the validity issues in the research process

3.0 MAIN CONTENT

3.1 Classification of Research Designs

A research design can be classified according to any of the following types:

3.1.1 Historical Research Design

This involves a systematic and objective enquiry into events, developments and experiences in the past. This is an ideal design for historians.

3.1.2 Survey Research Design

This is a research design in which a group of people or items is studied by collecting and analysing sample data or data from the entire population. If the survey involves study of a sample from the population, it is referred to as a **sample survey**. If it involves the study of the entire population, it is referred to as a **census survey.**

The idea of sampling is fundamental to survey research. A sound knowledge of sampling theory and techniques is necessary for the execution of a good survey research.

Survey research is classified into:

i. Procedure–based Survey including:
   - questionnaire survey
   - interview survey
   - observational survey
   - panel survey (where data are collected from a given sample at different time periods, and suitable for studying trends and fluctuations).

ii. Purpose–based

These are surveys classified by purposes they intend to accomplish and these include:

- Developmental Survey – which seeks to ascertain how some dimensions, variables and characteristics of a given population change with time.
- Descriptive survey – studies which aim at collecting data on, and describing in a systematic manner, the characteristics, features or facts about a given population.

Descriptive surveys do not require hypotheses, since they are merely concerned with a description of events as they occur.

- Correlation Survey – which seeks to establish what relationship exists between two or more variables.
- Public opinion survey – designed to find out the opinion of people in a given area towards an issue of interest.

3.1.3 Case Study Research Design

This involves intensive study geared towards a thorough understanding of a given social unit. It is worth noting that case studies are of limited generalisation. Only very few units are involved in case studies and as such, the findings cannot be generalised to the population.

3.1.4 Causal–Comparative or Ex – Post – Facto Research Design

This type of design seeks to establish cause – effect relationships. Here, the researcher attempts to link some already existing effect or observation to some variables as causative agents.

3.1.5 Experimental Research Design

This also establishes cause and effect relationships, except that it uses control groups.

3.2 Validity of a Research Design

Conclusions drawn from analysing survey data are only acceptable to the degree to which they are determined valid. Validity is used to determine whether research measures what it intended to measure and to approximate the truthfulness of the results. Researchers often use their own definition when it comes to what is considered valid. In quantitative research testing for validity and reliability is given. However some qualitative researchers have gone so far as to suggest that validity does not apply to their research even as they acknowledge the need for some qualifying checks or measures in their work. This is wrong. To disregard validity is to put the trustworthiness of your work in question and to call into question others confidence in its results. Even when qualitative measures are used in research they need to be looked at using measures of reliability and validity in order to sustain the trustworthiness of the results. Validity and reliability make the difference between "good" and "bad" research reports. Quality research depends on a commitment to testing and increasing the validity as well as the reliability of your research results.

Any research worth its salt is concerned with whether what is being measured is what is intended to be measured and considers the ways in which observations are influenced by the circumstances in which they are made. The basis of how our conclusions are made play an important role in addressing the broader substantive issues of any given study. For this reason we are going to look at various validity types that have been formulated as a part of legitimate research methodology.

3.2.1 Face Validity

This is the least scientific method of validity as it is not quantified using statistical methods. This is not validity in a technical sense of the term. It is concerned with whether it seems like we measure what we claim. Here we look at how valid a measure appears on the surface and make subjective judgments based on that. For example, if you give a survey that appears to be valid to the respondent and the questions are selected because they look valid. The researcher randomly asks a group of people, untrained observers if the questions appear valid to them. In research it's never sufficient to rely on face judgments alone and more quantifiable methods of validity are necessary in order to draw acceptable conclusions. There are many instruments of measurement to consider. So, face validity is useful in cases where you need to distinguish one approach over another. Face validity should never be trusted on its own merits.

3.2.2 Content Validity

This is also a subjective measure but unlike face validity we ask whether the content of a measure covers the full domain of the content. If a researcher wanted to measure introversion they would have to first decide what constitutes a relevant domain of content for that trait. This is considered a subjective form of measurement because it still relies on people's perception for measuring constructs that would otherwise be difficult to measure. Where it distinguishes itself is through its use of experts in the field or individuals belonging to a target population. This study can be made more objective through the use of rigorous statistical tests. For example you could have a content validity study that informs researchers how items used in a survey represent their content domain, how clear they are, and the extent to which they maintain the theoretical factor structure assessed by the factor analysis.

3.2.3 Construct Validity

A construct represents a collection of behaviours that are associated in a meaningful way to create an image or an idea invented for a research purpose. Depression is a construct that represents a personality trait which manifests itself in behaviours such as over sleeping, loss of appetite, difficulty concentrating, etc. The existence of a construct is manifest by observing the collection of related indicators. Any one sign may be associated with several constructs. A person with difficulty in concentrating may have A.D.D. but not depression. Construct validity is the degree to which inferences can be made from operationalisation (connecting concepts to observations) in your study of the constructs on which the operationalisation is based. To establish construct validity, you must first provide evidence that your data supports the theoretical structure. You must also show that you control the operationalisation of the construct, in other words, show that your theory has some correspondence with reality.

3.2.4 Internal Validity

This refers to the extent to which the independent variable can accurately be stated to produce the observed effect. If the effect of the dependent variable is only due to the independent variable(s) then internal validity is achieved. This is the degree to which a result can be manipulated.

3.2.5 Statistical Conclusion Validity

A determination of whether a relationship or co-variation exists between cause and effect variables requires ensuring adequate sampling procedures, appropriate statistical tests, and reliable measurement procedures. This is the degree to which a conclusion is credible or believable.

3.2.6 External Validity

This refers to the extent to which the results of a study can be generalised beyond the sample that is to say that you can apply your findings to other people and settings. Think of this as the degree to which a result can be generalised.

3.2.7 Criterion-Related Validity

This can also be referred to as instrumental validity. The accuracy of a measure is demonstrated by comparing it with a measure that has been demonstrated to be valid i.e. correlations with other measures that have known validity. For this to work you must know that the criterion has been measured well. And be aware that appropriate criteria do not always exist. What you are doing is checking the performance of your operationalisation against criteria. The criteria you use as a standard of judgment accounts for the different approaches you would use.

Predictive validity: Operationalisation's ability to predict what it is theoretically able to predict. The extent to which a measure predicts expected outcomes.

Concurrent validity: Operationalisation's ability to distinguish between groups it theoretically should be able to. This is where a test correlates well with a measure that has been previously validated.

When we look at validity in survey data we are asking whether the data represents what we think it should represent. We depend on the respondent's mind set and attitude in order to give us valid data. In other words we depend on them to answer all questions honestly and conscientiously. We also depend on whether they are able to answer the questions that we ask. When questions are asked that the respondent cannot comprehend or understand then the data does not tell us what we think it does.

SELF-ASSESSMENT EXERCISE

Outline the major activities necessary in a survey research.

4.0 CONCLUSION

This unit enumerates the different types of research and validity issues in the research process. The different types of research worth noting include:

(i) historical research;
(ii) survey research;
(iii) case study research;
(iv) causal-comparative research; and,
(v) experimental research.

The major types of validity enumerated are:

(i) face validity;
(ii) content validity;
(iii) construct validity;
(iv) internal validity;
(v) external validity;
(vi) statistical validity; and
(vii) criterion-related validity.

5.0 SUMMARY

You have been informed that research design can be thought of as the structure of research - it is the "glue" that holds all of the elements in a research project together. In a scientific research process, a design can be described using a concise notation that enables you to summarise a complex design structure efficiently.

In social and behavioural sciences, the most often used types of research designs are case study, survey, and causal-comparative. A survey research design is one in which a group of people or items is studied by collecting and analysing sample data or data from the entire population. If the survey involves study of a sample from the population, it is referred to as a **sample survey**. If it involves the study of the entire population, it is referred to as a **census survey.** Case studies involve intensive study geared towards a thorough understanding of a given social unit while causal-comparative research seeks to establish cause – effect relationships. Here, the researcher attempts to link some already existing effect or observation to some variables as causative agents.

You noted that conclusions drawn from analysing survey data are only acceptable to the degree to which they are determined valid. Validity is used to determine whether research measures what it intended to measure and to approximate the truthfulness of the results. Researchers often use their own definition when it comes to what is considered valid. In quantitative research testing for validity and reliability is given. However some qualitative researchers have gone so far as to suggest that validity does not apply to their research even as they acknowledge the need for some qualifying checks or measures in their work. This is wrong. To disregard validity is to put the trustworthiness of your work in question and to call into question others confidence in its results. Even when qualitative measures are used in research they need to be looked at using measures of reliability and validity in order to sustain the trustworthiness of the results. Validity and reliability make the difference between “good” and “bad” research reports. Quality research depends on a commitment to testing and increasing the validity as well as the reliability of your research results.

6.0 TUTOR-MARKED ASSIGNMENT

Discuss in detail what you would consider a valid research.

7.0 REFERENCES/FURTHER READING

Cooper, D. R. & Schindler, P. S. (2001). *Business Research Methods.* New York: McGraw-Hill.

Nworgu, B. G. (1991). *Educational Research: Basic Issues and Methodology*. Ibadan: Wisdom Publishers Ltd.

Onwe, O. J. (1998). *Elements of Project and Dissertation Writing: A Guide to Effective Dissertation Report.* Lagos: Impressed Publishers.

Osuala, E. C. (1993). *Introduction to Research Methodology*. Onitsha: Africana-Fep Publishers Ltd.

UNIT 2 COMPONENTS OF A RESEARCH DESIGN

CONTENTS

1.0 INTRODUCTION

In this unit, you will examine the different components of a research design. It suffices to say that the classic research design has four components: comparison, manipulation, control, and generalisation. It is on the basis of these components that you as a researcher can draw inferences concerning the criteria of causality and criterion of generalisation.

2.0 OBJECTIVES

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

- enumerate the basic components of a research design
- manage relevant research designs
- analyse research principles.

3.0 Main Content

3.1 The Components of a Classic Research Design

As noted in our introduction, a research design has four major components. These are:

(i) comparison,
(ii) manipulation,
(iii) control and
(iv) generalisation.

3.1.1 Comparison

The process of comparison focuses on the concept of co-variation and association between two or more variables. For example, a positive relationship exists between, say, the teaching method, X, and the achievement of students, Y; you will then think of finding a joint occurrence of both the teaching method (X) and a certain degree of achievement (Y). This implies that students are likely to achieve more after being exposed to the teaching method than before. Put differently, students who are studying under teaching method X will have higher achievement than those students who are studying under other teaching methods.

It follows that in order to assess the joint occurrence of the teaching method and achievement you need to make a comparison of the group of students exposed to the teaching method, X with the group of students that were not exposed to the method. You can also make a comparison of the group's achievement before and after exposing them to the teaching method, X. To measure co-variation therefore, the subject's scores on the dependent variable are evaluated before and after the introduction of the independent variable, or a group that is exposed to the independent variable is compared with the group that is not exposed. In the first case, a group is compared with itself; in the latter case, an experimental group is compared with a control group.

3.1.2 Manipulation

In quantitative methods, the term causality is meant to say that, if the variable Y is caused by the variable X, then an induced change in X will be followed by a change in Y. The assumption is that this relationship is asymmetrical, that is, one variable is the determining force, and the other is a determined response. If a teaching method, X, is to influence achievement, then you have to demonstrate that improvement in achievement will take place only after exposure to the teaching method. You can do this by some form of control over the introduction of the teaching method, so that you can measure achievement before and after the introduction of the teaching method. In an experimental setting, you can introduce the experimental stimulus yourself; in a natural setting, on the other hand, this may not be possible. In both natural and experimental (laboratory) settings, the major evidence required to determine the time sequence is that a change occurred only after the activation of the independent variable, in this case, the teaching method, X.

3.1.3 Control

Another criterion of causality requires that other factors be ruled out as rival explanations of the observed association between the variables under investigation. Such factors can likely invalidate the inference that the variables are causally related. This has been theoretically formulated as the problem of internal validity, which addresses itself to the question of whether the independent variable did in fact cause the observe response.

The factors jeopardising internal validity are often classified into those which are extrinsic to the research activity and those which are intrinsic and can impinge upon the results during the research period. More light can be thrown on these two classified factors.

Extrinsic factors

These factors refer to possible biases resulting from differential recruitment of subjects to the experimental and control groups in experimental research. These have been theoretically designated as selection factors that produce differences in the two comparison groups prior to the research activity.

Intrinsic factors

These refer to changes in the research subjects or in their backgrounds which occur during the period of research, or changes in the measuring instrument, or the reactive effect of the observation itself. The major intrinsic factors include:

1. **History**: History refers to all events occurring during the time of the research that are likely to affect the research subjects and provide a rival explanation for the change in the dependent variable. For example, in a study attempting to assess the effect of an election campaign on voters' behaviour, the hypothesis can be that propaganda to which voters are exposed during the campaign is likely to influence their voting behaviour. You will then compare the voting intentions of the subjects before and after exposure to the propaganda. The differences that you may discover in the voting intentions of the group that have been exposed to propaganda and those that have not could result from differential exposure to the material, or from events that occurred during this period.

2. **Maturation:** This is a second group of factors that may become plausible rival hypotheses. Maturation includes biological and psychological processes that produce changes in the subjects as time passes. These changes could influence the dependent variable and confound the research results. Suppose you want to evaluate the effect of a specific teaching method on student achievement and record the students' achievement before and after the teaching method has been introduced. Between the pre-test and the post-test, students may have become older and may be wiser. This change which is unrelated to the teaching method, could possibly explain the difference between the two tests. It has been noted that maturation, like history, constitutes a serious threat to the validity of causal inferences.

3. **Experimental mortality:** This refers to drop out problems that prevent the researcher from obtaining complete information on all cases or subjects. When subjects drop out selectively from the experimental or control group, the final sample on which complete information is available can be biased.

4. **Instrumentation:** Instrumentation designates changes in the measuring instruments between the pre-test and the post-test. To associate the difference between post-test and pre-test scores with the independent variable, you need to assume that repeated measurements with the same measuring instrument under constant conditions will yield the same result. If you cannot make such an assumption, observed differences can be attributed to the change in the measuring instrument and not necessarily to the independent variable. You can refer to the stability of measurement as reliability. The absence of stability can be a threat to the validity of experiments.

5. **Testing:** The process of testing may change the phenomenon being measured. The effect of being pre-tested might sensitise the subjects and improve their scoring on the post-test. A difference between post-test and pre-test scores could thus be attributed not necessarily to the experimental stimulus but rather to the experience gained by the subject while taking the pre-test.

The extrinsic and intrinsic factors that can threaten the internal validity of a design may be controlled by several operations. Control of intrinsic factors can be facilitated by the employment of a control group from which the experimental stimulus is withheld.

There are two methods of control that can be employed to counteract the effect of extrinsic factors. These methods include (i) matching; and, (ii) randomisation.

Matching: Matching involves equating the experimental and control groups on extrinsic variables that are presumed to be related to the research hypothesis. There are two suggested ways of matching groups:

(i) by precision or pair wise matching; and,
(ii) by frequency distribution. With precision matching, for each subject in the experimental group, another one with the same characteristics is selected for the control group. For example, to control the effect of age, for every individual in a specific age category in one group, there should be one in the same category in the second group. Having matched on the extrinsic variables, the investigator is assured that any difference found between the experimental and control groups cannot be due to the matched variables.

The main disadvantage of matching is however, the difficulty in matching a large number of variables. When there are many relevant characteristics that need to be controlled, it can be difficult to find matching pairs.

When matching by frequency distribution, you equate the experimental and control groups for each of the relevant variables separately rather than in combination. Instead of one-to-one matching, the two groups are matched on central characteristics. When matching for age, for example, the average age of one group should be equivalent to that of the second.

Randomisation: Randomisation is a process through which subjects are randomly assigned to the experimental and control groups. You can do randomisation either by flipping a coin to decide which subjects will be included in the experimental group; by using a table of random numbers; or by any other method which assures that any of the subjects has an equal probability of being assigned to either the experimental group or the control group.

3.2 Generalisation

Apart from internal validity, another significant research question concerns the generalisation of the research findings. External validity of research designs refers to the ability to generalise the research results. Two sources of external validity which can limit the generalisation of research findings have been discussed in the literature, including representativeness of the finding of the sample and the reactive arrangements in the research procedure.

Representativeness of the sample: The random assignment of subjects to experimental and control groups assures equality between the groups and thus contributes to the internal validity of a research. It does not necessarily assure representativeness of the population of study. Most results that prove to be internally valid might be specific to the sample selected for a given study. To enable generalisation beyond the limited scope of the specific study, you must be careful in selecting the sample using a sampling method that assures representation. Probability methods such as random sampling can make generalisations to larger and clearly defined populations possible.

Reactive arrangements: Results of a study should be generalised not only to a larger population but also to a real-life setting. When a study is carried out in a highly artificial situation, these generalisations cannot always be accomplished. In addition to the possible artificiality of an experimental setting, various features in the setting might be reactive and likely affect the external validity of the research results. For instance, a pre-test may influence the responsiveness of the subjects to the experimental stimulus; its observed effect would thus be specific to a population that has been pre-tested. The reactive effect of testing on the subjects can be avoided by carrying out a pre-test only, and the generalisation of the results can be improved by avoiding highly artificial situations.

SELF-ASSESSMENT EXERCISE

Present and discuss briefly the factors that can possibly jeopardise the internal validity of a research.

4.0 CONCLUSION

You have noted that a research design has four basic components among which are:

(i) Comparison;
(ii) manipulation;
(iii) control; and
(iv) generalisation.

The factors jeopardising internal validity of a research have been classified into those which are extrinsic to the research activity and those which are intrinsic and can impinge upon the results during the research period.

5.0 SUMMARY

A research design has four major components including:

(i) Comparison;
(ii) manipulation;
(iii) control; and
(iv) generalisation.

The process of comparison focuses on the concept of co-variation and association between two or more variables. The influence of research variables to others can be manipulated to cause and effect relationships. In quantitative methods, the term causality is meant to say that, if the variable Y is caused by the variable X, then an induced change in X will be followed by a change in Y. The assumption is that this relationship is asymmetrical, that is, one variable is the determining force, and the other is a determined response. Another criterion of causality has to do with control, which requires that other factors be ruled out as rival explanations of the observed association between the variables under investigation. Such factors can likely invalidate the inference that the variables are causally related.

The factors noted for jeopardising internal validity are often classified into those which are extrinsic to the research activity and those which are intrinsic and can impinge upon the results during the research period. Extrinsic factors refer to possible biases resulting from differential recruitment of subjects to the experimental and control groups in experimental research. Intrinsic factors refer to changes in the research subjects or in their backgrounds which occur during the period of research, or changes in the measuring instrument, or the reactive effect of the observation itself.

6.0 TUTOR-MARKED ASSIGNMENT

What are the components of a good research design and how can these components be used in assessing the validity of research findings.

7.0 REFERENCES/FURTHER READING

Cooper, D. R. & Schindler, P. S. (2001). *Business Research Methods.* New York: McGraw-Hill.

Nachimas, D. & Nachimas, C. (1976). *Research Methods in the Social Sciences.* New York: St Martin’s Press.

Nworgu, B. G. (1991). *Educational Research: Basic Issues and Methodology*. Ibadan: Wisdom Publishers Ltd.

Onwe, O. J. (1998). *Elements of Project and Dissertation Writing: A Guide to Effective Dissertation Report.* Lagos: Impressed Publishers.

Osuala, E. C. (1993). *Introduction to Research Methodology*. Onitsha: Africana-Fep Publishers Ltd.

MODULE 3 RESEARCH PROPOSALS

UNIT 1 THE RESEARCH PROPOSAL

CONTENTS

1.0 INTRODUCTION

This unit seeks to inform you about the purpose of a research proposal and how it can be used by you as a researcher and management decision maker. You will also learn about the different types of research proposals.

2.0 OBJECTIVES

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

- state the purpose of a research proposal
- enumerate the content and types of research proposals
- highlight the benefits of research proposal
- evaluate the quality of a research proposal.

3.0 MAIN CONTENT

3.1 Purpose of a Research Proposal

A proposal is generally looked at as an individual's or group of individuals' offer to render a service to a potential sponsor. The objectives of a business research proposal include:

(i) To present the management or research question to be answered through research and explain its importance
(ii) To discuss the research efforts of others who have worked on related management questions
(iii) To suggest the data necessary for solving the management question and suggest ways in which the data can be obtained, treated, and interpreted.

In presenting your research proposal, you should make your research plan clear and simple. You should also present your services and credentials in the best possible way to encourage the acceptance of your proposal among competing proposals. A proposal is often referred to as a work plan, prospectus, outline, statement of intent, or draft plan. The proposal tells you what, why, how, where, and to whom the research will be done. The benefits of the research must also be made clear to the sponsor.

Note that the more inexperienced a researcher is, the more important it is to have a well-planned and adequately documented proposal. The research proposal is essentially a road map, showing clearly the location from which a journey begins, the destination to be reached, and the method of getting there. A well-prepared research proposal would include potential problems that may be encountered along the way and methods for avoiding or working around them.

3.2 Benefits of the Research Proposal to the Sponsor

A research proposal allows the sponsor to assess the sincerity of the researcher's purpose, the clarity of his or her research design, the extent of his or her relevant background material, and fitness for undertaking the project. Depending on the type of research and the sponsor, various aspects of a standard proposal design are usually emphasised. The proposal displays the researcher's discipline, organisation, and logic. It therefore allows the sponsor of the research to assess both the researcher and the proposed research design, to compare them against competing proposals on current organisational, scholastic, or scientific needs, and to make the best selection for the project.

Comparison of the results of the research project is one of the first steps in the process of evaluating the overall research. By comparing the final product with the stated objectives, it will be easy for the sponsor to decide whether the research goal, that is, a better decision on the management question, has been achieved.

It has been observed that many managers, requesting research from an in-house, departmental research project, appear not to be adequately knowledgeable on the problem they are addressing. The research proposal acts as a catalyst for discussion between the person conducting the research and the manager. It is the role of the researcher to translate the management question, as described by the manager, into research question and to outline the objectives of the study. Upon review of the researcher's translation of the management question, the manager may discover that the interpretation of the problem does not reflect all the original symptoms. The proposal thus serves as the basis for additional discussion between the manager and the researcher till all aspects of the management question are clear and understood. It may be discovered that parts of the management question are not researchable, or not subject to empirical analysis. This then calls for an alternative design, such as qualitative or policy analysis study. The discussions will enable the sponsor and the researcher to agree on a carefully worded research question. In figure 1, you observe the proposal development process. The figure reveals that proposal development can work in an iterative way until the sponsor authorises the research to take off.

For most outside research contracts, the process appears to be different. In this case, you will submit the proposal in response to the request for bid or **request for proposal (RFP).** In this case, you will need to convince the sponsor that your approach to the research question differs from that indicated by the management question specified in the initial RFP. By so doing, you will show superior understanding of the management dilemma compared to other researchers with competing proposals.

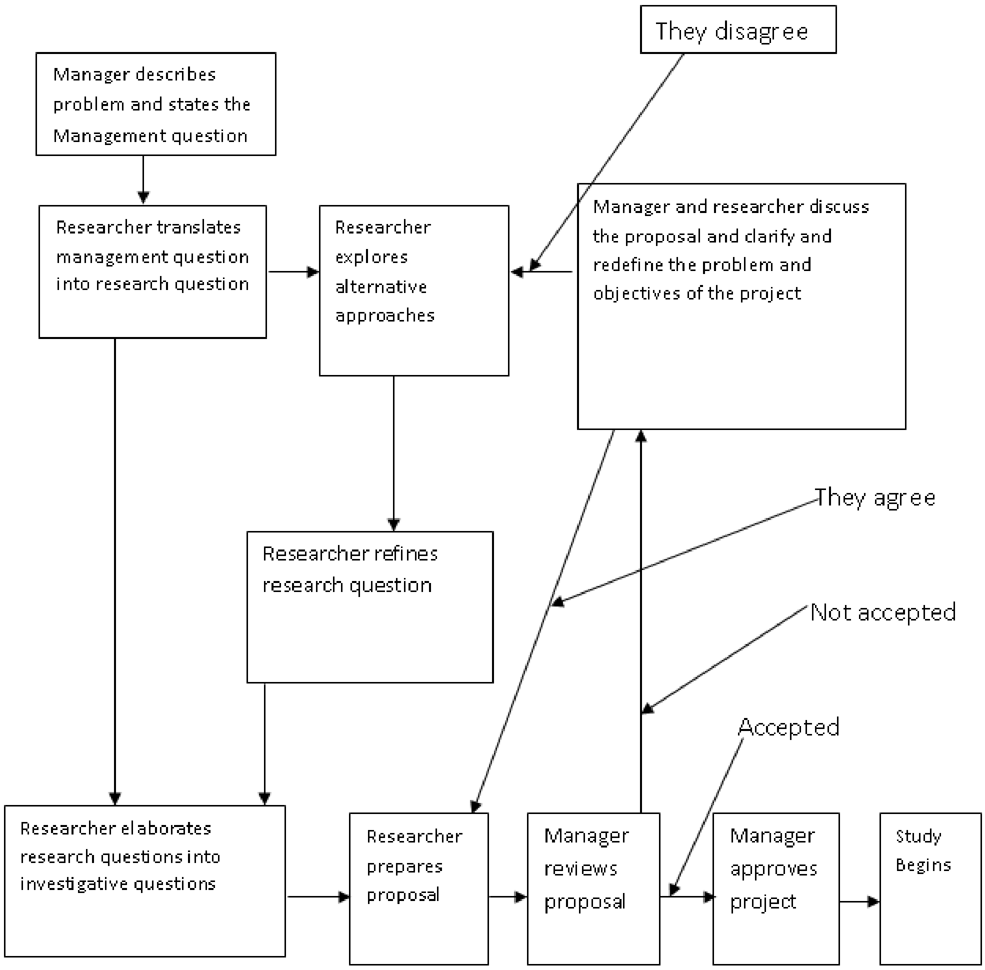


Fig. 1: The Proposal Development Process

3.3 Benefits of the Research Proposal to the Researcher

A proposal can be more beneficial to the researcher than to the sponsor. The act of writing a proposal encourages the researcher to plan and review the logical steps in the research project. The related management and research literature reviewed by the researcher in developing the research proposal prompts the researcher to assess previous approaches to similar management questions and revise the research plan accordingly. In addition, developing the proposal offers the researcher the opportunity to discover flaws in the logic, errors in assumptions, or even management questions that may not be adequately addressed by the objectives and design of the research project.

A thorough proposal process reveals all possible cost-related activities, thus allowing accuracy in cost estimates. Many of these cost-related activities are related to time, so that, a proposal benefits the researcher by forcing a time estimate for the project. The time and cost estimates encourage the researcher to plan the project so that work can progress steadily toward the deadline. Since human beings are often inclined to procrastinate, having a time schedule helps them work methodologically toward the completion of the research project.

You need to be guided by the fact that a poorly planned, poorly written, or poorly organised proposal can damage a researcher's reputation more than the decision not to submit a proposal.

3.4 Types of Research Proposals

Research proposals are often classified into those that are generated for internal audience and those generated for external audiences. An internal proposal is done by staff research specialists or by the research department of a given business organisation.

External proposals sponsored by university grant committees, government agencies, government contractors, non- profit organisations, or corporations can be further classified as either solicited or unsolicited. The larger the project, the more complex the proposal. In a public sector project, the complexity is generally greater than in a comparable private-sector project.

Three general levels of complexity have been noted:

(i) Exploratory studies
(ii) Small-scale studies
(iii) Large-scale studies.

An exploratory study generates the most simple research proposal. The more complex and common in business is the small-scale study. The large-scale professional study is the most complex proposal.

3.4.1 Internal Proposals

Internal proposals are regarded as being more succinct than external proposals. A three-page memo from the researcher to management, outlining the problem statement, study objectives, research design, and schedule is good enough to start an internal exploratory research. Privately and publicly operated businesses are usually concerned with how to solve a particular problem, make a decision, or improve an aspect of the business. Regardless of the intended audience, in the small-scale proposal, the literature review is not stressed and can be stated briefly in the research design. An executive summary is not mandatory for a small-scale internal research proposal. For funds to be committed however, it is necessary to provide time schedules and budgets for internal small-scale proposals.

3.4.2 External Proposals

An external proposal is either solicited or unsolicited. A solicited proposal is usually in response to a research need. Such proposal is likely competing against several other proposals for a grant. An unsolicited proposal represents a suggestion by a contract researcher for a research that might be done. An example is that of a consulting firm that proposes an omnibus study to a given trade association to address a problem arising from a change in the cultural or political-legal environment.

An unsolicited proposal has the advantage of not having competitors, but the disadvantage of having to speculate on the ramifications of a management dilemma facing an organisation's management. The writer of an unsolicited proposal must decide to whom the proposal should be sent, which can be a problem. Unsolicited proposals are time-sensitive, so that the window of opportunity is likely to close before a redirected proposal finds its appropriate recipient.

The most important parts of an external proposal are the objective, the design, qualifications, time schedule, and the budget. Note that in contract research, the results and objectives sections are the standards against which the completed project is measured. As the complexity of the project increases, more information will be required about the project management and the facilities and special resources.

Before leaving this unit, take a little time look at table 2. Compare the proposal modules that have been suggested for each type of study. This will most likely improve your understanding of research proposals.

Table 2: Proposal Modules: A Comparison of Management-Oriented Proposals and Student Proposals

Proposal Type/Proposal Modules	Management						Govern-ment	Student		
	Internal			External						
	Exploratory Study (ES)	Small-Scale Study (SS)	Large-Scale Study (LS)	ES	SS	LS	LS Contract	Term Paper	MS Thesis	Ph.D Thesis
Executive Summary		√	√	√	√	√	√			
Problem Statement	√	√	√	√	√	√	√	√	√	√
Research Objectives	√	√	√	√	√	√	√	√	√	√
Literature Review			√			√			√	√
Benefits of Study			√	√	√	√	√			√
Research Design	√	√	√	√	√	√	√		√	√
Data Analysis						√	√			√
Nature and form of Results		√	√		√	√	√		√	√
Qualification of Researchers				√	√	√	√			
Budget		√	√	√	√	√	√			
Schedule	√	√	√	√	√	√	√			√
Facilities and Special Resources			√	√	√	√	√		√	√
Project Management			√			√	√			
Bibliography			√			√	√	√	√	√
Appendices			√			√	√		√	√
Measurement			√			√	√			√

SELF-ASSESSMENT EXERCISE

Discuss briefly the qualities of an acceptable research proposal.

4.0 CONCLUSION

This unit has exposed you to the basic requirements of good research proposals. You learned about different types of proposals and their degrees of importance both to the research sponsor and the researcher.

You also learned about the process of a proposal development and a check list on proposal modules. The check list serves as a guide to the requirements of acceptable research proposals.

5.0 SUMMARY

A proposal has been defined as an individual's or group of individuals' offer to render a service to a potential sponsor. The objectives of a business research proposal were enumerated as follows:

(i) To present the management or research question to be answered through research and explain its importance
(ii) To discuss the research efforts of others who have worked on related management questions
(iii) To suggest the data necessary for solving the management question and suggest ways in which the data can be obtained, treated, and interpreted.

A research proposal allows the sponsor to assess the sincerity of the researcher's purpose, the clarity of his or her research design, the extent of his or her relevant background material, and fitness for undertaking the project. The research proposal displays the researcher's discipline, organisation, and logic.

Research proposals are often classified into those that are generated for internal audience and those generated for external audience. An internal proposal is done by staff research specialists or by the research department of a given business organisation.

External proposals sponsored by university grant committees, government agencies, government contractors, non -profit organisations, or corporations can be further classified as either solicited or unsolicited.

6.0 TUTOR-MARKED ASSIGNMENT

How does internal research proposal differ from an external research proposal?

7.0 REFERENCES FURTHER READING

Cooper, D. R. & Schindler, P. S. (2001). *Business Research Methods.* New York: McGraw-Hill.

Nworgu, B.G. (1991). *Educational Research: Basic Issues and Methodology.* Ibadan: Wisdom Publishers Ltd.

Onwe, O. J. (1998) *Elements of Project and Dissertation Writing: A Guide to Effective Dissertation Report.* Lagos: Impressed Publishers.

Osuala, E. C. (1993). *Introduction to Research Methodology.* Onitsha: Africana-Fep Publishers Ltd.

UNIT 2 STRUCTURING AND EVALUATING A RESEARCH PROPOSAL

CONTENTS

1.0 INTRODUCTION

The last unit presented you with the basic principles of research proposals. In this unit we introduce you to the best practice in structuring and evaluating a research proposal. The idea is to help you develop, write, and present an acceptable research proposal.

2.0 OBJECTIVES

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

- explain how an acceptable research proposal can be structured
- highlight the necessary parts of a research proposal
- describe how to impress a research sponsor and win a research grant.

3.0 MAIN CONTENT

3.1 Structuring the Research Proposal

Using table 2 as a reference point, you can design a set of modules that tailors your research proposals to the intended audience. Each of the following modules is flexible so that, the content and length may be adapted by you to specific needs. These modules include:

(i) Executive summary
(ii) Problem statement
(iii) Research objectives
(iv) Literature review
(v) Relevance/importance of the research
(vi) The research design and methodology
(vii) Data analysis
(viii) The nature and form of the results
(ix) Qualifications of the research crew
(x) The budget

3.1.1 The Executive Summary

The aim of executive summary is to allow a busy manager or sponsor to quickly understand the thrust of the research proposal. The summary represents essentially an informative abstract, giving executives the opportunity to grasp the essentials of the proposal without reading the details. Another aim of the executive summary is to secure a positive evaluation by the executive who passes the proposal on to his or her staff for a full evaluation. The executive summary should therefore include the following:

(i) Brief statement of the management dilemma and management question
(ii) The research objectives
(iii) The research questions
(iv) The benefits of your research and your approach.

If the proposal is unsolicited, a brief description of your qualifications will also be appropriate.

The research question appears to be at the heart of your case for research sponsor. The question that shapes what you present here would be: What is your core intellectual agenda? What is the research question that your proposal aims at answering and what is the significance of the research question? Is the research valuable to anyone? Who are the beneficiaries or ‘end users’ of the research?

3.1.2 Problem Statement

A well-thought problem statement will convince the sponsor to continue reading the proposal. You need to capture the reader's interest by stating the dilemma at hand, its background, its consequences, and the resulting management question. You need to emphasis the importance of answering the management question. You should also mention any restrictions or areas of the management question that will not be addressed.

You must be sure the problem statement is clear without the use of idioms. The problem statement should enable the sponsor understand the management dilemma and the question, its significance, and why something should be done to change the status quo.

3.1.3 Research Objectives

The research objectives should address the purpose of the investigation. They are aimed at laying out exactly what is being planned by the proposed research. In a descriptive type of research, the objectives can be stated as the research question. If the proposal is for a causal type of study, the objectives can be restated as hypothesis.

The objectives module should flow naturally from the problem statement so as to give the research sponsor specific, concrete, and achievable goals. The objectives should be listed either in order of importance or in general terms first, followed by specific terms (or research question followed by underlying investigative questions.

The research objectives are usually the basis for judging the rest of the proposal and, ultimately, the final report. You should verify the consistency of the proposal by checking to ensure that each objective is discussed in the research design, data analysis, and results modules or sections.

3.1.4 Literature Review

The literature review module examines the recent research studies, company data, or industry reports that act as basis for the proposed study. You need to begin your presentation of the related literature and relevant secondary data from a comprehensive perspective, moving to more specific studies that are associated with your problem statement. If the problem being addressed has a historical background, begin with the earliest references.

Try to avoid the extraneous details of the literature. A brief review of the information is enough. You must always refer to the original source of the literature information. If you discover something of interest in a quotation, find the original publication and ensure that you understand it. This will enable you avoid errors of interpretation or transcription. Try to emphasis the importance of results and conclusions of other studies, the relevant data and trends from previous research, and particular methods or designs that could be duplicated or should be avoided. You have to present how the literature applies to study you are proposing; you can show the weaknesses or faults in the design, presenting how you would avoid similar problems. If your proposal deals specifically with secondary data, discuss the relevance of the data and the bias or lack of bias associated with it.

You should end the literature module by scrutinising the important aspects of the literature and interpreting them in terms of the identified problem. Refine the problem as is needed in light of your findings from the literature review.

3.1.5 Relevance/Importance of the Research

This module describes the explicit benefits accruing from your research. The importance of doing the study now needs to be emphasised.

The module requires your understanding of what is most troubling to your sponsor. If it is a potential union unrest, for example, you cannot promise that an employee survey will prevent a union unrest. All you can do is to show the importance of the survey information, as well as its implications. This importance will more than likely allows the management to respond to employee concerns and initiate a linkage between those concerns and union unrest.

3.1.6 The Research Design and Methodology

At this point, you must have let the sponsor understand what the problem is, what your study goals are, and why it is important for you to carry out the study. You must have been also satisfied that you have made robust argument outlining the need for your particular research question to be answered, and you have articulated what is intellectually the best approach to seeking the answer. It is now time to present the design and methodology module. This describes what you are going to do in technical terms. This module should be made up of as many subsections as is needed to show the phases of the project. You should provide information on your proposed design for such tasks as sample selection and sample size, data collection method, instrumentation, procedures, and ethical requirements. You need to clearly explain what you propose to do and why you propose to do it.

You must set out clearly what you consider will be the best methodological approach to seeking the answers to your research question. Will your proposed methodology be primarily quantitative or qualitative? Will your research involve working in the field? When more than one methodology exists to approach the design, discuss the methods you have rejected and why the approach you selected is superior.

3.1.7 Data Analysis

This module involves a brief on the method used for analysing the data, especially for a large-scale contract research projects or doctoral thesis. If the project is a small project, you would include the proposed data analysis within the discussion of the research design. You need to describe the way you will handle the data and the theoretical basis for using your selected techniques. This presentation will assure the sponsor that you are following correct assumptions as well as using theoretically sound procedure for data analysis.

The data analysis module is so important to evaluating contract research proposals that the researcher should contact an expert to review the latest techniques available for use in the particular research project and compare this to your proposed techniques.

3.1.8 The Nature and Form of the Results

This module will enable the sponsor to go back to the statement of the management question and research objectives and perhaps discover that each of the research goals has been covered. In this module, you are required to specify the types of data to be obtained and the interpretations that will be made in the analysis. If the data are to be released to the sponsor for proprietary reasons, you should make sure this is reflected.

The module also contains the contractual statement telling the sponsor the types of information to be received. Statistical conclusions, applied findings, recommendations, action plans, models, strategic plans, and the like are examples of what we refer to as forms of results.

3.1.9 Qualifications of the Research Crew

This module should begin with the principal investigator followed by similar information on all individuals involved with the project. There are two critical elements here:

1. Professional research competence including relevant research experience, the highest academic degree held, and membership in business and technical societies.
2. Relevant management experience.

Giving so many competing individual researchers, research specialty firms, and general consultancies providing research services, the sponsor will like to be assured that the researcher(s) is or are professionally competent. Past research experience has been noted as the best barometer of competence, followed by the highest academic degree earned. In documenting relevant research experience, you as a researcher should present a concise description of similar projects that you undertook. Highest degree usually follows your name, for example, Onyemaechi Onwe, Ph.D in Economics. Society or association memberships provide some evidence that the researcher is knowledgeable on the latest methodologies and techniques. Membership in the West African Research and Innovation Management Association (WARIMA) is a good one.

Highly established and well-known research institutes and companies do sometimes subcontract specific research activities to firms or individuals that specialise or offer specific resources or facilities. This is especially the case for studies involving quantitative research techniques such as econometric techniques of data analysis, or those involving qualitative techniques such as in-depth personal interviews and focus groups. You can provide brief profiles of these institutes or companies in this module only if you believe their inclusion will enhance your research credibility.

3.1.10 The Budget

You need to present the budget in the form requested by the sponsor. The first and foremost thing to know is your sponsor or funder. Read the funder’s guidelines on what may or may not be included in costing of a project. Some sponsors will allow equipment, other will not. Some will allow a certain percentage of the budget to be claimed as indirect costs and others will not. Some sponsors require secretarial assistance to be individually budgeted for. Others insist it should be included in the research director’s fees or the operation’s overhead.

An effective budget should typically not be more than two pages. Table 3 below is an example of a format that can be used for a small contract research project. It is advisable to put additional information, backup details, and hourly time and payment calculations into an appendix if required or kept in your file for future reference.

The budget statement in an internal research proposal will be based on employee and overhead costs. Budget presented by an external research organisation is not just the wages and salaries of its employees but sometimes the person-hour price that the contracting firm charges.

External research firms avoid giving detailed budget for fear of the possibility of disclosing their costing practices publicly. They fear this would reduce their flexibility in negotiating contracts.

It is extremely important that you try to retain all information you used in generating the research budget. If you used quotes from external contractors, get the quotation in writing. If you estimated time for interviews, keep explicit notes on how the estimate was made.

Table 3: Sample Research Proposal Budget

Budget Items	**Rate**	**Total Days**	**Charge**
A. Salaries:			
1. Research Director, Joseph Onwe	N2000/day	20 days	N40,000
2. Associate	N1000/day	20 days	N20,000
3. Research Assistants (2)	N200/day	60 days	N12,000
4. Secretarial (1)	100/day	20 days	N2,000
		Sub-Total	**N74,000**
B. Other Costs:			
5. Employee Services & Benefits			-
6. Travel			N10,000
7. Office Supplies			N2,500
8. Telephone			N4,000
9. Rent			-
10. Other Equipment			-
11. Publications and storage Costs			N1,500
		Sub-Total	**N18,000**
C. Total of Direct Costs			**N92,000**

D. Overhead Support			**N20,000**
		TOTAL FUNDING REQUESTED	**N112,000**

3.1.11 Schedule

A research schedule needs to include the major research phases, their time table, and their milestones signifying completion of a phase. Major phases, for example, may be as follows:

Phase 1: Exploratory interview
Phase 2: Final research proposal
Phase 3: Questionnaire revision
Phase 4: Field interviews
Phase 5: Editing and coding
Phase 6: Data analysis
Phase 7: Report generalisation

You should have an estimated time schedule for each of these phases. You need to have also the people assigned to complete each phase.

If the research project is large and somehow complex, it will be helpful for you and your sponsor to chart your schedule using the **critical part method (CPM)**. As you can observe from Figure 2 below, in a CPM chart, the nodes represent major milestones and the arrows suggest the activity needed to get to the milestone. More than one arrow pointing to a node will indicate all those tasks that must be completed before the milestone can be reached. A number is usually placed along the arrow showing the number of days, weeks or months required for that task to be completed. The pathway from start to the end that takes the longest time to complete is referred to as the **critical path**, because any delay in an activity along that path will delay the completion or end of the entire project.

Note that software programmes that are designed for project management are available for installation in personal computers and laptops. Your ICT department can help.

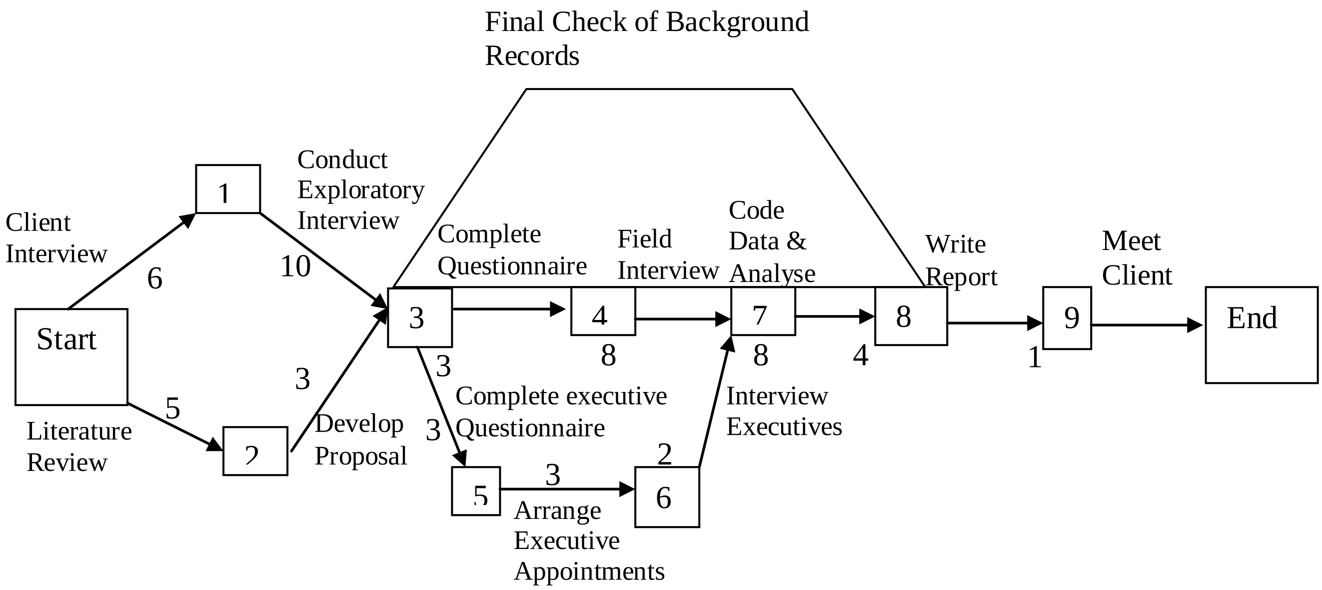


Fig. 2: The Critical Path Method (CPM) Schedule

Summary
The Milestones:
3 Proposal approval
7 Interview completed
9 Final report
Critical Path:
Start – 1 – 3 – 4 – 7 – 8 – 9 – ***End***
Completion Time: 6 + 10 + 3 + 8 + 8 + 4 + 1 = 40 working days

3.1.12 Facilities and Special Resources

Research projects often require special resources and facilities. You need to describe these in detail. For instance, a contract exploratory study may require specialised facilities for focus group sessions. Computer-assisted telephone or other interviewing facilities may be needed. Your proposed data analysis may require sophisticated computer algorithms, and therefore, requiring your access to an adequate computer system. These requirements vary depending on the nature and complexity of the study.

The proposal should therefore, contain the list of relevant facilities and resources that you will use for effectiveness. The costs for such facilities should be detailed in your budget.

3.1.13 Project Management

The aim of this module is to prove to the sponsor that the research team is organised in a way to carry out the project effectively and efficiently. For complex research projects, you would require a master plan, showing how all the phases will be brought together. Suggested inclusions of the master plan are:

1. The research team's organisation
2. Management procedures and controls for executing the research plan
3. Examples of management and technical reports
4. The research team's relationship with the sponsor
5. Financial and legal responsibility
6. Management competence

It is most helpful for you to use tables and charts in presenting the master plan. If several researchers are part of the team, you need to indicate the relationships between the researchers and their assistants. Sponsors are interested in knowing that the research director is an individual capable of leading the team and acting as a useful liaison to the sponsor.

Note in addition that procedures for information processing, record control and expense control are critical to large operations. You need to show these as part of the management procedures.

You need to record the type and frequency of progress reports. The sponsor would like to be kept up-to-date. You need also to delineate the sponsor's limits to control during the research process.

In this module, you should also discuss any details such as printing facilities, clerical help, or information-processing capabilities to be provided by the sponsor, rather than you as the researcher.

Payment frequency and timing should be covered by you in the master plan. You need to provide proof of financial responsibility and overall management competence in the master plan.

3.1.14 Bibliography and Appendices

A bibliography is necessary for all research projects requiring a literature review. You must use the bibliography format required by the sponsor. If the sponsor did not specify any, you should consult a standard style manual which will give you the necessary details for preparing the bibliography.

You need to provide an appendix, in the form of glossary of terms, when you have many words that are unique to your particular research topic and are not easily understood by the general management community.

This glossary should consist of terms and definitions. Any acronyms used by you should be defined.

The appendix should also contain samples of your proposed measurement instruments. This will enable the sponsor to be able to discuss particular changes in one or more of the instruments. You can also include any detail that reinforces the body of the proposal. This may include your curriculum vitae, profiles of firms or individuals to which the project may be subcontracted, budget details, and extended descriptions of special facilities or resources.

3.2 Evaluating the Research Process

Research proposals are usually subject to formal or informal reviews. Formal reviews are regularly done for solicited proposals. The formal review process typically includes:

1. Development of review criteria, using the request for proposal (RFP) guidelines
2. Assignment of points on each criterion, using a universal scale
3. Assignment of a weight for each criterion, based on the importance of each criterion
4. Generation of a score for each proposal, representing the sum of all weighted criterion scores.

It is the sponsor who assigns the criteria, the weights, and the scale to be used for scoring each criterion before receiving the proposals. The proposal is evaluated with this checklist of criteria. Points reflecting the sponsor's assessment of how well the proposal meets the company's needs are recorded for each criterion. Points such as 1 through 10, with 10 being the highest number of points assigned to the best proposal for a particular criterion. After the review, the weighted criterion scores are then added to get a cumulative total. It is the proposal with the highest number of points that will win the research contract.

It is worth noting that, in practice, many factors do contribute to the acceptance and funding of a research proposal. Beyond the required modules as discussed above, other factors can quickly eliminate a research proposal from consideration or improve the sponsor's reception of the proposal. Among these factors are:

1. Neatness
2. Organisation, in terms of being both logical and easily understood
3. Completeness in fulfilling the request for proposal's (RFP's) specifications, including budget and time schedule

4. Appropriateness of writing style
5. Submission within the RFP's timeline.

It is important to stress the importance of the technical writing style. The writing style must be in such a way that the sponsor understands clearly the problem statement, the research design, and the methodology. The sponsor must understand why the proposed research should be funded and the exact goals and concrete results that is expected from the research.

You should also endavour to make sure the research proposal meets the specified RFP guidelines, including budgetary restrictions, and schedule deadlines. If your schedule does not meet the expected deadline, your proposal will be disqualified. If your budget is too high relative to competing those of competing proposals, it will be rejected. However, a low budget compared to those of competing proposals may suggest that something is missing or there is something wrong with you as a researcher or your research team.

Finally, a late proposal will not be reviewed under normal circumstances. Lateness communicates a level of disrespect for the sponsor. A late proposal can also communicate a weakness in project management, which can raise an issue of professional competence.

SELF-ASSESSMENT EXERCISE

Enumerate and discuss briefly the major phases of a research.

4.0 CONCLUSION

You have been informed on the different necessary modules of a research proposal, ranging from the executive summary to the bibliography and appendices. Emphasis was placed on scheduling as this is the most critical part of a proposal's preparation. For an easy and practical scheduling, you were introduced to the critical path method (CPM). You were also informed that research proposals are subject to both formal and informal reviews. Formal reviews are regularly done for solicited proposals.

5.0 SUMMARY

This unit informs you that in structuring a research proposal, you can design a set of modules that tailors your research proposal to the intended audience. The following modules have been suggested.

(i) Executive summary
(ii) Problem statement
(iii) Research objectives
(iv) Literature review
(v) Relevance/importance of the research
(vi) The Research design and methodology
(vii) Data analysis
(viii) The Nature and form of the results
(ix) Qualifications of the research crew
(x) The budget.

Research proposals are usually subject to formal or informal reviews. Formal reviews are regularly done for solicited proposals. The formal review process typically includes:

1. Development of review criteria, using the request for proposal (RFP) guidelines
2. Assignment of points on each criterion, using a universal scale.
3. Assignment of a weight for each criterion, based on the importance of each criterion
4. Generation of a score for each proposal, representing the sum of all weighted criterion scores.

6.0 TUTOR-MARKED ASSIGNMENT

Discuss the importance of an executive summary needed for a research proposal.

7.0 REFERENCES/FURTHER READING

Cooper, D. R. & Schindler, P. S. (2001). *Business Research Methods.* (New York: McGraw-Hill.

Nworgu, B.G. (1991). *Educational Research: Basic Issues and Methodology.* Ibadan: Wisdom Publishers Ltd.

Onwe, O. J. (2009). "Acceptability of Research Proposals in Social Sciences: Important Considerations in the West African Sub-Region." *Journal of Social Policy and Society,* vol. 4, (1).

Onwe, O. J. (1998). *Elements of Project and Dissertation Writing: A Guide to Effective Dissertation Report.* Lagos: Impressed Publishers.

Osuala, E. C. (1993). *Introduction to Research Methodology.* Onitsha: Africana-Fep Publishers Ltd.

MODULE 4 THE RESEARCH HYPOTHESES

UNIT 1 RESEARCH QUESTIONS AND HYPOTHESES

CONTENTS

1.0 INTRODUCTION

Research questions are mostly aimed at ascertaining the reason or reasons for the existence of the problem or problems under investigation. They attempt to uncover the basic questions to be answered in the course of the research process. Research questions are usually categorised into broad and specific questions. Answers to these questions provide the basis for solving identified research problems and testing the research hypotheses. The research hypotheses are postulations or statements about what the researcher thinks about expected behaviours which are subject to confirmation through statistical facts. This unit discusses at length the processes in the design of research questionnaire from research questions, statement of hypotheses, and hypotheses testing.

2.0 OBJECTIVES

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

- design questionnaire for survey studies
- administer research questions and achieve high rate of returns for your research questionnaires
- construct different hypothetical statements
- highlight the different types of research hypotheses and the tools used in testing them.

3.0 MAIN CONTENT

3.1 Statement of the Research Question and Questionnaire Design

3.1.1 Statement of the Research Question

The main consideration in the statement of research questions is the specific research problem being addressed. For example, if the specific research problem is: 'Distress in the Nigerian Banking Industry,' two research questions may be:

(i) What is the origin of bank distress in Nigeria?
(ii) What is the role of deregulation on the incidence of bank distress in Nigeria?

3.1.2 Construction of Questionnaire

Questionnaires are derived from the research questions and are the most frequently used instruments for the collection of primary data or information, especially in survey research. Every question in questionnaire must produce responses that will help answer the stated research questions.

Before constructing a questionnaire, you need to identify clearly the objective of the questionnaire. You must know what information should be obtained from the respondents, using the questionnaire to be constructed.

The important factors that can guide you in the construction of questionnaires include:

(i) The characteristics of the Sample Involved - Knowledge of such characteristics as the level of education, socio–economic class, age, sex, and the like will help you design appropriate questionnaire.

(ii) The type of questionnaire format to adopt – You as a researcher should know whether the questions should be of fixed–response type or open–ended as explained below.

Structured or fixed-response questionnaire: This is the type of questionnaire in which respondents are given response alternatives by the researcher.

Unstructured or open–ended questionnaire: In this type of questionnaire, the researcher does not provide the response alternatives. Respondents are free to give their individual responses

You need to bear in mind that the longer the questionnaire, the lower will be the response rate and vice versa. The length of questionnaire should therefore be minimised if your aim is to achieve a high response rate.

You may also be guided by the following:

i. Avoid questions which do not relate to the research objectives, research questions and research hypotheses.
ii. Avoid such leading questions as those questions, beginning with such terms as "In view of the fact that."
iii. Avoid unnecessary presumptions about the respondents.

The format to be adopted will depend on the dept of the information required.

3.1.3 Validation of the Constructed Questionnaire

A commonly used validation process is to send the constructed questionnaire to a panel of experts in the area of the research interest or problem for validation. The questionnaire to be sent to the experts should be accompanied by clear guidelines on what is expected of them. These guidelines should include the purpose of the study, the research questions, and the research hypotheses.

After validating the content of the questionnaire, the next step is to do a pilot test of the questions on a sample of the intended respondents. This will enable you ascertain how the respondents will likely react with the questionnaire in terms of clarity, ambiguity and coverage.

3.1.4 Administration of the Questionnaire

Three traditional modes of administrating the questionnaire include:

(i) Personal Interview: Involves a face–to–face interview whereby you as the researcher present the questions to the respondents with the aim of helping in the clarification of the questions. This has the highest response rate.

(ii) Telephone Interview:_ Involves phone calls to the respondents. This mode is not very feasible, especially in developing countries with clumsy telephone services.

(iii) Mail: This is the least costly mode. It however has the least rate of response.

3.1.5 Characteristics of a Good Questionnaire

For a research questionnaire to provide information needed to meet the objectives of the research, it must have the following characteristics:

(i) Relevance
(ii) Consistency
(iii) Usability
(iv) Clarity
(v) Quantifiable
(vi) Legibility.

3.2 Statement of Hypotheses

The statement of hypotheses is an important part of a research proposal, especially in social sciences and humanities. The testable form of a hypothesis is stated in two forms: the null hypothesis: (i) the null hypothesis; and, (ii) the alternate hypothesis. Null hypotheses appear as negative non-parametric statements, while alternate hypothesis appear as positive non-parametric statements. A null hypothesis is the hypothesis which states that 'no difference' or 'no relationship' exists between two or more variables. It is often referred to as hypothesis of 'no effect' or 'no difference.' An alternate hypothesis is a hypothesis that specifies any of the possible conditions not anticipated in the null hypothesis. It specifies conditions which will hold if the null hypothesis does not hold.

For example, consider the following null and alternate non-parametric hypothesis:

Null hypothesis, Ho: There is no significant difference between the smoking habits of a black man and a white man.

Alternate hypothesis, Ha: There is a significant difference between the smoking habits of a black man and a white man.

In parametric terms, null and alternate hypotheses can be illustrated in the following forms, but not necessarily in negative or positive forms as in the case of non-parametric hypotheses:

Ho: $\mu = 0$, implying that the population average or mean is equal to zero.

Ha: $\mu > 0$, implying that the population average is greater than zero.

3.3 Types of Hypotheses

There are several ways of classifying hypotheses but we will be interested in classifying them as either research hypotheses or statistical hypotheses.

3.3.1 Research (or Non-Parametric) Hypotheses

Research hypotheses are postulations about the relationships between two or more variables that are highly important in solutions to the problem of interest in your research. Research hypotheses may take any of the following forms:

(i) The use of minimum deposit requirements will facilitate real banking in Nigeria.
(ii) Nigerian workers' poor attitude to work is due to the fact that supervisors do not have powers to discipline them.

Observe that research hypotheses do not express the variables in quantitative or measurable terms and therefore, they cannot be directly tested by statistical methods.

3.3.2 Statistical (or Parametric) Hypotheses

A statistical hypothesis is a proposition about population parameters, such as the population mean or the population standard deviation, which is to be verified on the basis of the data or information obtained from a sample of the population. Statistical hypotheses express the relationship between two or more variables in statistical, quantitative, or measurable terms. The statistical parameter on which the test will be based is specified and the variables are reduced to numerical quantities. This is the form in which these hypotheses are tested, unlike the research hypotheses.

An hypothesis provides you with the necessary guide in your search for the solution to the problem being addressed in the research proposal. Hypothesis helps you ensure that you do not waste time and energy in searching for the solutions to the problem anywhere and anyhow.

Formulation and use of appropriate hypotheses is crucial and fundamental to the success of your research activities. This is essentially the case in studies for which you are seeking for cause and effect relationships.

You need to know the situations which do not require your formulation of hypotheses. The situations are in cases of preliminary or exploratory investigations that are aimed at gaining more information about some happenings, or where you have no knowledge about possible solutions to the problem under investigation.

In using hypotheses, you should endavour to make your hypotheses comprehensive enough to cover all aspects of your problem focus. It has been noted that the use of hypotheses may lead to researchers neglecting or ignoring important aspects of findings from the research data, not covered by the hypotheses. This only happens when the hypotheses is not appropriate and comprehensive enough.

3.4 Qualities of Good Hypotheses

A good hypothesis should be:

1. Testable. It should be in such a way that empirical evidence relating to its validity can be obtained.
2. A statement of an expected relationship between two or more variables. The aim of using hypothesis is usually to find out how two or more variables are related. A good hypothesis should therefore, specify the expected relationship between the variables of interest in measurable terms.
3. Plausible. Hypotheses should be based on what is consistent with reasoning. It should not only relate to the problem of interest but also its likelihood as a solution to the problem should not be in doubt.
4. Consistent with current knowledge. A hypothesis should not contradict established knowledge. If it does, it will appear unreasonable.
5. Unambiguous. A hypothesis needs to be stated in clear unambiguous and simple terms. Any term you use in formulating the hypothesis should be operationally defined by you.

3.5 Sources of Hypotheses

As discussed briefly earlier, research hypotheses can be obtained from the following sources:

1. Experience: what you know about the problem under investigation.
2. The literature: from the review of related literature, you can obtain useful ideas relating to possible solutions to the problem under investigation
3. Theory: from theories you can derive hypotheses through the process of deductive reasoning, for example, "If A is true then B will be true."
4. Previous findings: findings of previous studies can also serve as sources of hypotheses. Such findings may give rise to some new issues which need to be resolved. Such issues then form the basis for formulating pertinent hypotheses. Also, deductions made from the findings of previous studies can be useful in formulating hypotheses.

SELF-ASSESSMENT EXERCISE

Discuss briefly the importance of statement of hypothesis in scientific research.

4.0 CONCLUSION

This unit introduces you to the concept of hypothesis and how it can be formulated, the statement of research questions, questionnaire design, types of hypotheses, qualities of good hypothesis, and sources of hypotheses. You were informed that research hypotheses can be obtained from four major sources, namely:

(i) experience;
(ii) literature;
(iii) theory;
(iv) previous research findings.

The emphasis was on statistical or parametric hypotheses.

5.0 SUMMARY

Hypothesis can be defined as a conjectural proposition, an informed, intelligent guess about solution to a problem. It is an assumption or proposition whose veracity and validity must be established.

A hypothesis provides you with the necessary guide in your search for the solution to the problem being addressed in the research proposal and helps you ensure that you do not waste time and energy in searching for the solutions to the problem anywhere and anyhow.

There are several ways of classifying hypotheses but we will be interested in classifying them as either research hypotheses or statistical hypotheses.

Research hypotheses may take any of the following forms:

(i) The use of minimum deposit requirements will facilitate real banking in Nigeria.
(ii) Nigerian workers' poor attitude to work is due to the fact that supervisors do not have powers to discipline them.

A statistical hypothesis is a proposition about population parameters, such as the population mean or the population standard deviation, which is to be verified on the basis of the data or information obtained from a sample of the population.

6.0 TUTOR-MARKED ASSIGNMENT

Data on the age of individual customers were observed as: 15 years, 16 years, 18 years, 17 years, and 22 years.

(a) What is the average age of customers according to the data?
(b) Using a 1% significance level, test the hypothesis:

Ho: U = 17.5 years
Ha: U < 17.5 years

7.0 REFERENCES/FURTHER READING

Nworgu, B.G. (1991). *Educational Research: Basic Issues and Methodology.* Ibadan: Wisdom Publishers Ltd.

Onwe, O. J. (2007). *Statistical Methods for Business and Economic Decisions: A Practical Approach*. Lagos: Samalice Press.

Onwe, O. J. (1998). *Elements of Project and Dissertation Writing: A Guide to Effective Dissertation Report.* Lagos: Impressed Publishers.

Osuala, E. C. (1993). *Introduction to Research Methodology.* Onitsha: Africana-Fep Publishers Ltd.

UNIT 2 HYPOTHESIS TESTING I

CONTENTS

1.0 INTRODUCTION

The term hypothesis was discussed in detail in the last unit. In this unit and the one that follows, we will discuss the testing tools available for you. In testing hypotheses, assumptions about the population parameter, such as the population average or mean, are made in advance, and the relevant population sample provides the information needed for the test of the assumptions. This unit looks at hypotheses testing by focusing on such concepts as the sampling distribution, null and alternate (or research) hypotheses, the level of significance, rejection values, and decision rules.

The first step in testing hypothesis is to formulate the hypothesis in statistical terms. For example, if the purpose of your investigation is to establish that educated individuals have higher income than do uneducated individuals, the statistical hypothesis might be that there is a positive correlation between education and income, or that the average income of highly educated group will be larger than the average income of a group with lower level of education. In both of these cases, the statistical hypothesis is formulated in terms of descriptive statistics (such as correlation or average), as well as a set of specifying conditions about these statistics (such as positive correlation or difference between the means).

2.0 OBJECTIVES

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

- explain what hypothesis is all about
- distinguish between parametric and non-parametric hypotheses

- enumerate the tools used in testing parametric hypotheses
- state and test parametric hypotheses

3.0 MAIN CONTENT

The tools used in testing statistical hypothesis are often referred to as parametric tools. Those used in testing non-statistical hypothesis are called the non-parametric tools. In the following few examples we discuss how these tools can be used in testing hypothesis.

3.1 The Parametric Tools

The two basic parametric tools for testing statistical or parametric hypothesis to be discussed here are the student t-statistic and the z-statistic.

3.1.1 The Student t-Statistic

The student t-statistic is used in testing hypotheses concerning the population mean or average, especially in cases involving a relatively small sample size; ($n \leq 30$). Its application is better illustrated by an example. Before any illustration, it will be appropriate to review the process of hypotheses testing as it concerns the student t-statistic.

The Decision Values of t

The aim of any hypothesis testing is to either accept or reject a given null hypothesis. The decision to either accept or reject any hypothesis is based on two values of the test statistic, in this case the t-statistic. These two values are referred to as the decision values, they are:

1. The critical or rejection value
2. The calculated or statistical value.

1. The Critical or Rejection Value of t

The critical value of t is obtained from the t distribution table, with known level of significance, α , and the number of degrees freedom, n-1, where n represents the number of observations.

Assume n=15 observations. With 5% level of significance and n-1 = 15-1 = 14 degrees of freedom obtained from the t-distribution table as t=1.76, for one-sided rejection region or t=2.145, for two-sided rejection regions (see figure 3).

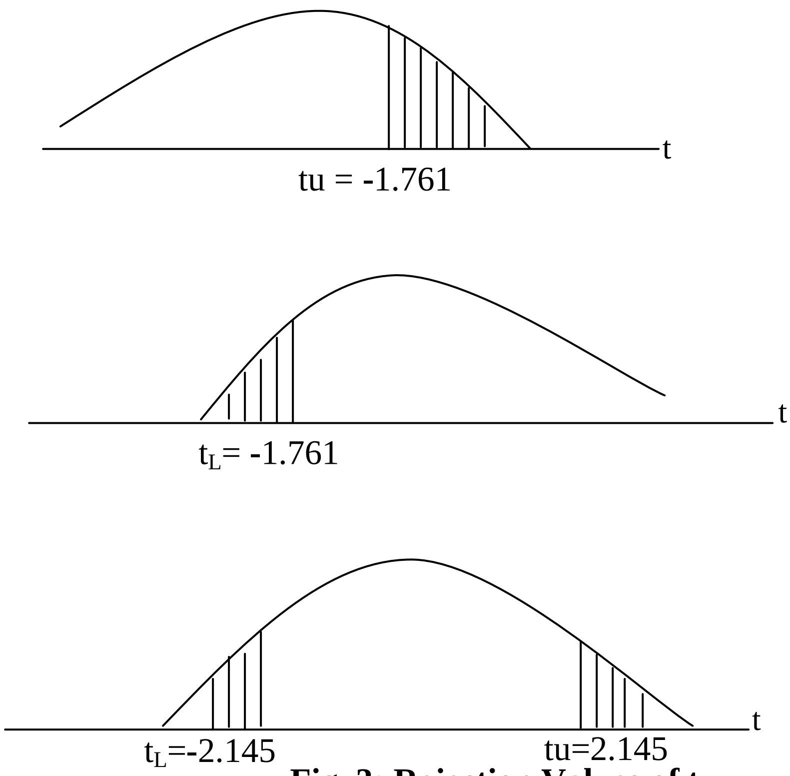


Fig. 3: Rejection Values of t

2. The Calculated Value of t (t_c)

This is the value obtained using the statistical information needed for testing the stated hypotheses. To calculate this value, we simply apply the formula:

$$tc = \frac{\bar{X}-\mu_0}{S_{\bar{X}}}$$

Where $\bar{X}$ = sample mean

μ_0 = the hypothesised population mean

S_X = the standard error of the sample mean.

$$\bar{S}_X = \frac{S}{\sqrt{n-1}}$$

Recall that S = Sample Standard deviation.

Having obtained the two values of the test-statistic, we compare them to know whether or not the null hypothesis should be rejected.
If tc > tu, reject Ho (the null hypothesis)
If tc < tu, accept Ho(the null hypothesis)
And
If tc < tL, reject Ho(the null hypothesis)
If tc > tL, accept Ho(the null hypothesis)
Where tc, tu, and tL refer to calculated, upper, and lower critical values of t respectively.

Example

Upon examining the monthly billing records of a mail-order book company, the auditor takes a sample of 10 of its unpaid accounts. The accounts receivable were:

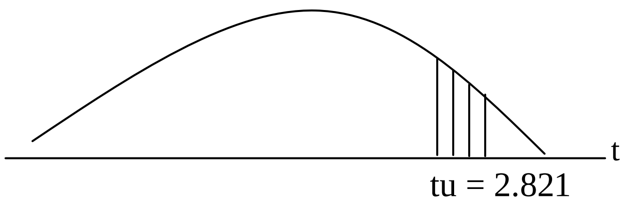
₦4, ₦5, ₦7, ₦7, ₦9, ₦10, ₦11, ₦12, ₦18, ₦33.

Based on the observed accounts, the auditor hypothesizes that on the average, the accounts receivable is greater than N15. Following this hypothetical belief, we want to test at 1 percent level of significance, the hypotheses:

Ho: μ = N15 (null hypothesis)
HA: μ > N15 (Alternate hypothesis)

Solution

First the inequality sign (>) indicates the use of one-sided, upper rejection region in the decision of either to reject or accept the hypothesis. It follows that, using the table on t-distribution, the critical or rejection value of t is as indicated below, with $\alpha = 0.01$ and
n-1 = 10-1 = 9 degrees of freedom



The calculated value of t is obtained as follows:

$$tc = \frac{\bar{X} - \mu o}{S_{\bar{X}}}$$

Where $\bar{X} = \frac{\Sigma X}{n} = \frac{166}{10} = 11.6$

$\mu_o = 15$

$$S_{\bar{X}} = \frac{S}{\sqrt{n}}$$

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

X		X^2
4		16
5		25
7		49
7	49	
9		81
10	100	
11	121	
12	144	
18	324	
33		1089
$\sum X = 116$		$\sum X^2 = 1998$

$$\frac{(\sum X)^2}{n} = \frac{(116)^2}{10} = \frac{13456}{10} = 1345.6$$

$\sum X^2 = 1998$

$$S = \sqrt{\frac{1998-1345.6}{9}} = \sqrt{72.49}$$

$= 8.51$

Thus, $S_{\bar{X}} = \frac{S}{\sqrt{n}} = \frac{8.51}{\sqrt{10}} = 0.851$

It follows that,

$$tc = \frac{\bar{X}-\mu o}{S_{\bar{X}}}$$

$$= \frac{11.6-15}{0.851} = -3.995$$

Decision

Since tc (-3.995) < tu(2.82), we accept the null hypotheses (Ho = ₦15), and conclude that on the average, the account receivable is not significantly greater than ₦15.

3.1.2 The Z-Statistic

Like the t-statistic, z-statistic is a parametric tool used in testing parametric hypotheses. In this discussion, we are interested in the use of z-statistic in testing hypotheses that involve percentages or proportion of subjects having particular responses to particular research question or issues.

3.1.2.1 Tests of Hypotheses for Sample Proportions: One Sample

For large samples, the applicable test-statistic for sample proportions is the z-statistic. The rejection value of z for a given level of significant is obtained from the z-distribution table. And the calculated value of z for tests of proportions can be obtained by:

$$Z_C = \frac{P_S - P}{\sigma_{pS}}$$

Where p_S = the sample proportion or percentage

$$= \frac{\text{Number of successes in a sample}}{\text{Sample size}}$$

P = Population proportion as indicated by the null hypothesis, Ho.

σ_{ps} = standard error of the sample proportion

$$= \frac{\sqrt{P(1-P)}}{n}$$

n = Sample size

Example

Consider a supermarket that sells packaged men's shirts. The management learns from past experience that 15% of all shirts sold are returned to the supermarket by customers who complain that the shirts do not fit properly. In an attempt to correct this situation, the manufacturer of the shirts redesigned them and finds that, of the next 500 sales, 60 shirts were returned.

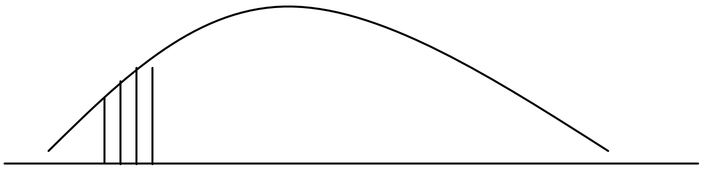
Problem is to test, at 5% level of significance, to see if there has been a significant decrease in the population proportion of returns.

Solution

We are required to test,

Ho: $P = 0.15$
Ha: $P < 0.15$ $\alpha = 0.05$

Observe that the alternative hypothesis, Ha, calls for a one-sided test, using the lower rejection region. The rejection value of Z at $\alpha = 0.05$ is –1.64 (from the Z-distribution table) as indicated by figure 4 below.



Z_L = -1.64

Fig. 4: Rejection Value for Z (α = 0.05)

The calculated value of Z can be obtained as:

$$Zc = \frac{P_S - P}{\sigma_{pS}}$$

where $ps = \frac{60}{500} = 0.12$ = proportion of returns

$P = 0.15$ (from the null hypothesis, Ho)

$$\sigma_{ps} = \sqrt{\frac{P(1 - P)}{n}} = \sqrt{\frac{0.15(1 - 0.15)}{500}}$$

$$= \sqrt{\frac{0.15(0.85)}{500}} = 0.016$$

It follows that,

$$Z = \frac{0.12 - 0.15}{0.016} = -1.875$$

Decision

Since Z(-1.875) < Z_L(-1.64), we reject Ho and conclude that there has been a significant decrease in the population proportion of returns as a result of the changes made in the design of the shirts.

SELF-ASSESSMENT EXERCISE

Present briefly the decision rules in testing parametric hypotheses, using the student t-statistic.

4.0 CONCLUSION

This unit discussed hypotheses testing by focusing on such concepts as the sampling distribution, null and alternate (or research) hypotheses, the level of significance, rejection values, and decision rules. It suggests that the first step in testing hypothesis is to formulate the hypothesis in statistical terms. The emphases were on the student t-statistic and the Z-statistic as a tool in testing parametric hypotheses.

5.0 SUMMARY

The tools used in testing statistical hypothesis are often referred to as parametric tools. Those used in testing non-statistical hypothesis are called the non-parametric tools. The first example of parametric tools is the student t-statistic. The student t-statistic is used in testing hypotheses concerning the population mean or average, especially in cases involving a relatively small sample size; ($n \leq 30$). Like the t-statistic, z-statistic is a parametric tool used in testing parametric hypotheses. The discussions concentrated on the use of z-statistic in testing hypotheses that involves percentages or proportion of subjects having particular responses to particular research question or issues.
The aim of any hypothesis testing is to either accept or reject a given null hypothesis and, the decision to either accept or reject any hypothesis is based on two values of the test statistic, in this case the t-statistic. These two values are referred to as the decision values, namely: (i) The critical or rejection value; and, (ii) the calculated or statistical value.

6.0 TUTOR-MARKED ASSIGNMENT

XYZ Manufacturing Company pays 10 different employees the following salaries in Naira.

Monthly Salary (Y): 12,000, 12,600, 14,000, 13,000, 15,500, 17,000, 16,000, 16,800, 18,000, 19,500.

Test at 1% level of significance, the hypothesis that the average monthly salary of employees at XYZ Manufacturing Company is greater than 16,000 naira.

7.0 REFERENCES/FURTHER READING

Nworgu, B.G. (1991). *Educational Research: Basic Issues and Methodology.* Ibadan: Wisdom Publishers Ltd.

Onwe, O. J. (2007). *Statistical Methods for Business and Economic Decisions: A Practical Approach.* Lagos: Samalice Press.

Onwe, O. J. (2009). "Acceptability of Research Proposals in Social Sciences: Important Considerations in the West African Sub-Region." *Journal of Social Policy and Society,* vol. 4, (1).

Onwe, O. J. (1998). *Elements of Project and Dissertation Writing: A Guide to Effective Dissertation Report.* Lagos: Impressed Publishers.

Osuala, E. C. (1993) *Introduction to Research Methodology*. Onitsha: Africana-Fep Publishers Ltd.

UNIT 3 HYPOTHESIS TESTING II

CONTENTS

1.0 INTRODUCTION

This unit is a continuation of the discussions of hypothesis testing. In the last unit, the student t-statistic used in testing parametric of statistical hypothesis was discussed. In this unit, the other statistic used in testing parametric hypothesis for a large sample size, the Z-statistic will be presented and discussed. The unit will end with non-parametric tools used in testing non-parametric hypotheses.

2.0 OBJECTIVES

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

- explain various tools used in testing hypothesis
- differentiate between parametric and non-parametric hypotheses
- analyse hypothesis testing.

3.0 MAIN CONTENT

3.1 The Non-Parametric Tools and Testing Procedures

There exist certain phenomena or variables in business statistics which can hardly be described quantitatively. The mode of gathering information on these variables requires the use of nominal and ordinal scales which do not meet standard requirements of parametric statistics. The valid inferential statistical tests for these types of variables are non-parametric tests.

Non-parametric test procedures involve either:

1. Those procedures whose test-statistic does not depend upon the form of the underlying population distribution from which the sample data were drawn;
   or
2. Those procedures which are not concerned with the population parameters;
   or
3. Those procedures for which the data are of little strength to warrant meaningful arithmetic operations.

Conditions under which non-parametric statistics can_be used are outlined as follows:

1. When the hypothesis to be tested does not involve a population parameter.
2. When there are no assumptions of normality about the distribution of the variables.
3. When data are gathered from such weak measuring scales as ranking, frequency counts, and some subjective measuring scales.
4. When results are needed fast and no statistical sophistication is required.

There are two most commonly used non-parametric statistical methods in business statistics, these are:

1. Chi-square (χ^2) statistic
2. Spearman rank correction (r_s)

At this level of discussion, however, we shall examine the Chi-square method.

3.1.1 The Chi-Square Method

The Chi-square test can only indicate whether or not a set of observed frequencies differ significantly from the corresponding set of expected frequencies and not the direction in which they differ. In practice, there are two types of Chi-square (χ^2) tests:

1. Test of goodness-of-fit
2. Test of independence and /or homogeneity.

The test of goodness-of-fit

The test of goodness-of-fit is employed in situations whereby the researcher's objective is to find out whether or not a set of observed frequencies fits closely the theoretical or expected frequencies. The applicable formula for the test of goodness-of-fit is:

$$\chi^2_c = \sum_{i=1}^{r} \frac{(f_{oi} - f_{ei})^2}{fei}$$

Where f_o = observed frequencies
f_e = expected frequencies
χ^2_c = calculated value of χ^2

The number of degrees of freedom for r number of row entries is r-1 for the test of goodness-of-fit. Given the number of degrees of freedom and the level of significance, the rejection value of χ^2 can be obtained from the χ^2-distribution table.

Consider the following practical example:
In a study to determine customer preferences among two banking services: current account and savings account, 120 customers were asked to respond to the question,
"Please indicate which service you prefer most."

Current Account _______ ? Savings Account _______ ?

The responses indicate that 65 of the respondents preferred current account while 55 preferred savings account. This data can be summarised thus:

Table 4: Number Preferring Banking Services

BankBanking Service	Num Number preferring
Cc Current Account	65
SsSavings Account	55
Total	120

We want to test the following non-parametric hypotheses:

$H_{O:}$ The customers do not show any preference for either Current Account or Savings Account.

$H_{A:}$ The customers show some preference for either Current Account or Savings Account.

We test the hypotheses at 1% level of significance (that is, $\alpha = 0.01$). so that the rejection value, with r-1 = 2-1 = 1 degree of freedom, where r is the number of row categories, that is, Current Account and Savings

Account, is 6.635 from the χ^2 -distribution table. This rejection value is illustrated in figure 5 below.

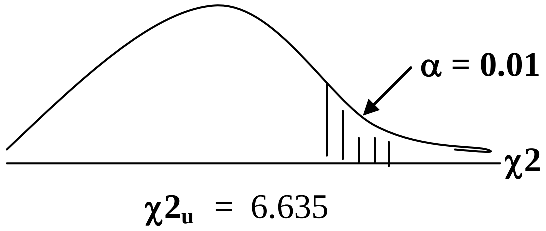


Fig. 5: Rejection Value for χ^2 (d.f = 1, α = 0.01)

The calculated value of χ^2

Bank Banking Service	Num No. preferring (f_{oi})	Exp Expected Preference (f_{ei})
Curr Current Account	65 65	60 60
SaviSavings Account	55 55	60 60
Total	120 120	120 120

Note: Expected frequency (f_{ei}) = $\frac{\text{Number of Respondents}}{\text{Number of rows}}$

$$= \frac{120}{2} = 60$$

It follows that:

$$\chi^2 c = \Sigma \frac{(f_{oi}-f_{ei})^2}{f_{ei}}$$

$$= \frac{(65-60)^2}{60} + \frac{(55-60)^2}{60}$$

$$= 0.42 + 0.42$$

$$= 0.84$$

Decision

Since χ^2_c (0.84) < χ^2_u(6.635), we do not reject Ho, implying that the customers do not show any preference for either Current or Savings Account. The observed pattern of 65 preferring Current Account and 55 preferring Savings Account is not statistically significant.

Test of independence and homogeneity

A test of independence aims at ascertaining whether two or more variables are dependent upon each other while tests of homogeneity aim at ascertaining whether the characteristics of two or more population variables are the same.
Two variables are said to be independent (or not associated with each other) if the distribution of one is not related to the distribution of the other. Chi-square (χ^2) test of independence can therefore be used to test if the distributions of two variables in a population are independent of each other.

Two important assumptions of the Chi-square test of independence are worth mentioning:

1. The relevant data are randomly drawn from a population of interest.
2. Two criteria are used in the cross-classification of the observations, and each observation must belong to only one criterion. This cross-classification gives rise to what is referred to as an n-contingency table.

A contingency table is the table in which the observed and expected frequencies associated with the various levels of two variables are presented. The table is named by the number of rows(r) and number of columns (c) it has, as an (r x c) contingency table. If the table has 2 rows and 3 columns, it will be referred to as a (2 X 3) contingency table. Note that in practice, the expected frequency is recorded in the same cell as the observed frequency. The expected frequency is, however, differentiated from the observed frequency by enclosing it in a bracket inside the cell.

As an example, assume that the director of Enugu State Chamber of Commerce is interested (for planning purposes) in learning more about the international participants in its annual international trade fairs. From the local hotel association, a list of past participants is obtained. The director plans to send questionnaires to the participants on the list to find out why they participated in the trade fairs, how much they spent, how long they stayed, and what their future plans are. The director speculates that an offer to send each respondent a free gift will increase the rate of return of questionnaire responses. To test this proposition, questionnaires were mailed to a random sample of 30 persons with the offer of the free gift. Questionnaires were also mailed to another random sample of 30 persons with no gift offer. The results are shown in the following (2 X 2) contingency table:

Table 5: A contingency Table of Numbers Responding to Questionnaire

	Questionnaire		
Gift	Returned	Not Returned	Total
Offered	22	8	30
Not Offered	14	16	30
Total	**36**	**24**	**60**

The director's proposition can formally be hypothesised as follows:

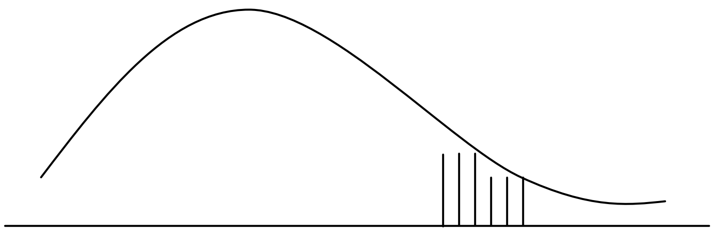
Ho: The population proposition of questionnaire returns is independent of the promise of a gift.

Ha: The population proportion of questionnaire returns is dependent upon the promise of a gift.

We want to test these hypotheses at 5% level of significant ($\alpha = 0.05$) to see if there is a significant difference in the proportion of returns when the gift is offered.

Solution

From χ^2-distribution table, we observe the rejection value of χ^2, with $(r-1)(c-1) = (2-1)(2-1) = 1$ degree of freedom, at $\alpha = 0.05$ to be $\chi^2u = 3.84$ This can be illustrated as in figure 6 below.



$\chi 2_u(3.84)$

Fig. 6: Rejection Value for χ^2 (df= 1; α = 0.05)

For tests of independence and homogeneity, the calculated value of Chi-square (χ^2) is obtained by:

$$\chi^2 c = {}^r\Sigma \; {}^c\Sigma \frac{(f_{ij}-f_{ij})^2}{f_{ij}}$$

Where the expected frequencies, f_{ij} can be obtained by the definition:

$$f_{ij} = \frac{R_{it}C_{jt}}{N}$$

Where, R_{it} = row total
C_{jt} = column total
N = the grand total

From the above table 5, the totals are

$R_{1t} = 30$; $R_{2t} = 30$; $C_{1t} = 36$
$C_{2t} = 24$; $N = 60$

It follows that the expected frequencies, f_{ij}, will be:

$$f_{11} = \frac{R_{1t}C_{1t}}{N} = \frac{(30)(36)}{60} = 18$$

$$f_{12} = \frac{R_{1t}C_{2t}}{N} = \frac{(30)(24)}{60} = 12$$

$$f_{21} = \frac{R_{2t}C_{1t}}{N} = \frac{(30)(36)}{60} = 18$$

$$f_{22} = \frac{R_{2t}C_{2t}}{N} = \frac{(30)(24)}{60} = 12$$

Note in our formulations that, f_{ij} = expected frequency of observation in the ith row and jth column, so that:

f_{11} = Expected frequency in the 1st row and 1st column
f_{12} = Expected frequency in the 1st row and 2nd column
f_{21} = Expected frequency in the 2nd row and 1st column
f_{22} = Expected frequency in the 2nd row and 2nd column

As mentioned earlier, the expected frequencies in our example can be presented along with the corresponding observed frequencies in the (2X2) contingency table as follows:

	Questionnaire		
Gift	Returned	Not Returned	Total
Offered	22 (18)	8 (18)	30
Not offered	14 (18)	16 (12)	30
Total	36	24	60

The figures in brackets are the corresponding expected frequencies.
The calculated $\chi 2$ can now be presented as follows:

$$\chi^2 = \frac{(f_{ij} - f_{ij})^2}{f_{ij}}$$

$$= \frac{(22 - 18)^2}{18} + \frac{(8 - 12)^2}{12} + \frac{(14 - 18)^2}{18} + \frac{(16 - 12)^2}{12}$$

$$= 0.89 + 1.33 + 0.89 + 1.33 = 4.44$$

Decision

Since $\chi 2_c(4.44) > \chi 2_u(3.84)$, we reject the null hypothesis, Ho, and infer that the population proportion of questionnaire returns in our example is dependent upon the promise of a gift.

SELF-ASSESSMENT EXERCISE

Explain the major difference between a test of homogeneity and a test of independence.

4.0 CONCLUSION

This unit has further exposed you to the way hypotheses are tested using the relevant tools. Of major significance is the use of non-parametric test tools involving tests of homogeneity and independence. This was discussed at length using a simple practical example.

5.0 SUMMARY

Non-parametric test procedures involve either:

1. Those procedures whose test-statistic does not depend upon the form of the underlying population distribution from which the sample data were drawn;

   or

2. Those procedures which are not concerned with the population parameters;

   or

3. Those procedures for which the data are of little strength to warrant meaningful arithmetic operations.

There are two most commonly used non-parametric statistical methods in business statistics, including:

1. Chi-square (χ^2) statistic
2. Spearman Rank Correction (r_s)

At this level of discussion, however, we shall examine the Chi-square method.

In practice, there are two types of Chi-square (χ^2) tests:

1. Test of goodness-of-fit
2. Test of independence and /or homogeneity.

6.0 TUTOR-MARKED ASSIGNMENT

A sample survey of users of public libraries was conducted to investigate the reading habits of men and women. The results are as follows:

	Types of literature preferred		
	Fiction	Non – fiction	Total
Men	132	102	234
Women	168	98	266
Total	**300**	**200**	**500**

Test at five percent level to see if there is any evidence that women show greater preference for fiction than men.

7.0 REFERENCES/FURTHER READING

Nworgu, B.G. (1991). *Educational Research: Basic Issues and Methodology*. Ibadan: Wisdom Publishers Ltd.

Onwe, O. J. (2007). *Statistical Methods for Business and Economic Decisions: A Practical Approach*. Lagos: Samalice Press.

Onwe, O. J. (2009). "Acceptability of Research Proposals in Social Sciences: Important Considerations in the West African Sub-Region." *Journal of Social Policy and Society,* vol. 4 (1).

Onwe, O. J. (1998). *Elements of Project and Dissertation Writing: A Guide to Effective Dissertation Report.* Lagos: Impressed Publishers.

Osuala, E. C. (1993). *Introduction to Research Methodology*. Onitsha: Africana-Fep Publishers Ltd.

MODULE 5 DATA PRESENTATION AND ANALYSIS

Unit 1 Data, Sources of Data, and Methods of Data Collection
Unit 2 Data Presentation and Analysis

UNIT 1 DATA, SOURCES OF DATA, AND METHODS OF DATA COLLECTION

CONTENTS

1.0 INTRODUCTION

This unit contains a detailed discussion of data as the basic component of research. Without data, you cannot think of research on business, economic, social and health problems. The medical doctor needs data on patient's illness in order to ascertain the root cause of a given illness. Economists need data to make predictions and proffer solutions to economic problems. The lawyer needs data to effectively defend his or her subject in the court of law. You can now appreciate the importance of a unit devoted primarily to data, their sources and methods of collection.

2.0 OBJECTIVES

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

- explain classifications of data
- enumerate sources of data
- highlight different techniques in the collection of research data.

3.0 MAIN CONTENT

3.1 Data and Sources of Data

3.1.1 Data

The term data can be referred to as unbiased set of statistic observations or information. There are basically two types of data.

1. Qualitative data involve random variables that yield categorised responses.
2. Quantitative data involving random variables that yield numerical responses.

Below is a summary of how the two types of data can be generated.

Data Type		Question Type	Responses
Qualitative	→	Do you own company shares	Yes --- No ---
Quantitative:			
Discrete	→	How many stick of cigarette do you smoke a day?	Number ------
Continuous	→	How tall are you?	Number ------

Fig. 7: Types of Data

Note that **discrete quantitative** data are numerical responses which arise from a counting process, while **continuous quantitative** data are numerical responses which arise from a measuring process.

3.1.2 Sources of Data

Sources of data include:

1. Publications: These include published data from such government agencies as the National Bureau of Statistics (NBS), Nigeria Deposit Insurance Corporation (NDIC), Central Bank of Nigeria (CBN) Statistical Bulletins, the World Bank and the World Tables. These are the major sources of secondary data.
2. Experimentation: Data can be obtained through experiments.
3. Survey: With appropriate questionnaire instrument, reliable data can be obtained through survey. Survey is the major source of primary data.

3.2 Methods of Data Collection

There are five major methods of data collection.

1. Observation: This involves monitoring the situation under investigation by trained observers.
2. Inspection: Inspection requires test of objects (e.g. test of weight, tensile strength, blood pressure, etc).
3. Abstraction from records: Analysis of past records (or secondary data).
4. Questionnaire: Data can be collected through responses to structured and unstructured questionnaires.
5. Interviewing: This involves a person-to-person administration of questionnaires.

Data collection is an activity requiring statistical inquiry. It is highly beneficial to you to examine the stages of statistical inquiry as presented below.

3.2.1 Stages of Statistical Inquiry

The stages of statistical inquiry are as follows:

Stage 1: Problem statement

A statistical inquiry must begin with explicitly identity and definition of the problem of interest. The inquiry should not be launched in general terms. As a case in point, an investigation into the problem of labour turnover may not equally affect all parts of the firm concerned. Inquiry may not be necessary in many departments where there is no problem of labour turnover. The problem area may be a particular department, or a particular process or product. A careful statement of the problem of interest will give the researcher terms of reference from which relevance data can be collected for analysis.

Stage 2: Decision on the best approach

In line with the terms of reference in stage 1, you need to determine how you will approach the problem. Statistical evidence may already be available from past research. Another person may have already faced the same problem; it will be a waste of time to repeat an investigation. Many good inquiries start with a literature review in which one can read the published materials available on the problem of interest. Other inquiries begin with a thorough survey of all past records available in house. For example, the investigation of the labour turnover problem in a particular department might begin with examination of the in-house personnel files. The basic question would be: Why did employees in this department leave? Did the explanations these people gave fit into a pattern of behaviour that identifies the cause of the problem in focus?

Was it working conditions, level and type of remuneration, supervisory problems, or what?

Stage 3: Definition of the extent of the inquiry

You need to state the extent of the inquiry. Shall it extend to the whole population? Should it be just a sample survey?

Stage 4: Determination of the instrument for the inquiry

You need to know whether questionnaire should be used as an instrument for the inquiry. A pilot inquiry may be necessary to test out the questionnaire in order to find out if there are unforeseen ambiguities in the questions.

Stage 5: Data collection

This is the stage whereby observations are made and recorded. Data can be collected in several ways including field observations and survey (for primary data), and journals and publications (for secondary data). Any data collection method can also be used to gather information concerning different hypotheses.

Most statistical inquiries are in the form of interviews. It is essential to appoint interviewers, brief them adequately and ensure that they conduct the interviews in a proper manner. Any conclusion or inference drawn from a badly conducted series of interviews will be meaningless. Other inquiries do not involve interviews. In these cases, data are collected by enumerators who record facts as they become available.

Stage 6: Editing and Classification of the Data

The outcome of stage 5 will be a mass of raw data, in a very indigestible form. Most of the time, these need editing. The data are then tallied, tabulated and summarised in a useful form; in a form that they could be easily analysed.

Stage 7: Data analysis

At this stage, you are now in the position to analyse the data. The data analysis needs the use of those who are knowledgeable about the subject of interest and skilled in statistical methods. At this stage, all available data is analysed according to research objectives, research questions, and hypotheses. The analytical method to be used will depend basically on the type of research as well as the research design.

Stage 8: Data presentation and report writing

The outcome of any survey should be a set of proposals that would remedy the identified problem. In making these proposals, it you are required to present the data in a simple and convincing manner, as part of the report to the appropriate authority. Such report must quote the terms of reference of the inquiry. You should use such tools as tables, charts, diagrams to show the findings of the survey.

Your report should suggest the causes or reasons for the identified problem, and make authentic recommendations for the solutions. Statistical surveys are generally expensive, and the researcher will be judged by his or her cost effectiveness in demonstrating the true facts and recommending the most likely cure in the circumstances.

SELF-ASSESSMENT EXERCISE

Explain why it is necessary to begin statistical inquiries with the problem identification.

4.0 SUMMARY

This unit has exposed you to the basic principles of data collection. To make data collection an easy process, the unit suggests some data sources and stages of statistical inquiry, ranging from problem identification to data presentation and report writing.

5.0 SUMMARY

The term data has been referred to as unbiased set of statistic observations or information. There are basically two types of_data: (i) qualitative data involving random variables that yield categorised responses; and, (ii) quantitative data involving random variables that yield numerical responses.

Data can be sourced at three different levels: (i) publications; (ii) experimentation; and, (iii) survey. The specific ways in which data can be obtained at these levels were presented.

6.0 TUTOR-MARKED ASSIGNMENT

The ABC Transport Company has 100 similar buses. Management suspects that its control of these buses is not adequate. What suggestions could you make for investigating the use of the buses, and tightening managerial control over the buses?

7.0 REFERENCES/FURTHER READING

Nachimas, D. & Nachimas, C. (1976). *Research Methods in the Social Sciences.* New York: St Martin's Press.

Nworgu, B. G. (1991). *Educational Research: Basic Issues and Methodology.* Ibadan: Wisdom Publishers Ltd.

Onwe, O. J. (1998). *Elements of Project and Dissertation Writing: A Guide to Effective Dissertation Report.* Lagos: Impressed Publishers.

Onwe, O. J. (2007). *Statistical Methods for Business and Economic Decisions: A Practical Approach.* Lagos: Samalice Press.

Onwe, O. J. (2009.) "Acceptability of Research Proposals in Social Sciences: Important Considerations in the West African Sub-Region." *Journal of Social Policy and Society*, vol. 4(1).

Osuala, E. C. (1993). *Introduction to Research Methodology*. Onitsha: Africana-Fep Publishers Ltd.

UNIT 2 DATA PRESENTATION AND ANALYSIS

CONTENTS

1.0 INTRODUCTION

In this unit, you will discover the basic principles and tools of data presentation in business statistics. There are some basic things to be considered before presenting and analysing a research data. It is the aim of this unit to look at such things.

2.0 OBJECTIVES

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

- explain the first things to do before presenting and analysing a research data
- enumerate the acceptable tools for presentation and analysis of data
- analyse data for a given research.

3.0 MAIN CONTENT

The preliminaries in data presentation and analysis are two-fold. The first is data editing and, the second is the choice of the appropriate tool of analysis.

3.1 Editing

Before a given data can be presented for analysis and interpretation, it must be edited and coded. By editing, we mean the examination of the given data in order to detect errors that may cause inconsistency if they are used for analysis in their original form. Through editing, these errors can be corrected accordingly.
There are two types of editing:

1. Field editing
2. Central editing

3.1.1 Field Editing

Field editing is a process whereby the researcher makes his or her records complete and correct without adding subjective information to his or her sources. It involves the presentation of collected information in a readable form such that all information gathered is properly reported.

3.1.2 Central Editing

Central editing ensures maximum consistency in information by correcting any inconsistency in the collected data, which might create problems in the analysis and interpretation of the results. There are four possible errors that should be watched out for in central editing:

1. **Arithmetic or numeric errors**: These are errors that involve the wrong recording of the units in responses. Information may be wrongly reported in months when they are requested to be in years.
2. **Errors of transposition:** This error occurs when a response is entered in the wrong place. For example, a question that asks for the respondent's state of origin may generate an answer about the ethnic group to which the respondent belongs.
3. **Errors of inappropriate response:** These errors occur when a respondent gives a relevant response but not in the exact form that is required.
4. **Errors of omission:** These errors are difficult to edit and, in most cases, are interpreted to mean "no response".

3.2 Coding

Coding enables the researcher group responses into limited number of classes or categories for ease of analysis. A set of rules are observed when grouping the responses into classes or categories. The rules include:

1. **Exclusiveness:** This requires that a data item or response must be placed in one cell of a given category set. This is essentially relevant in a situation where a respondent or response fits into several categories.
2. **Exhaustiveness:** This is a requirement that all data categories or cells must be able to provide the necessary data for answering the research questions and testing the research hypotheses.

3.3 Data Analysis

Research is generally meant to generate data for analysis, and this can result in a large amount of statistical information mostly in its raw stage. For the generated data to be useful in attaining the objectives of a research, they have to be reduced to manageable dimensions through analysis.

Two types of data analysis are in place:

1. Descriptive analysis
2. Causal analysis

3.3.1 Descriptive analysis

Descriptive analysis deal with the study of such research variables as profiles of the respondents, organisations, groups or any other subjects. Descriptive analysis may be either qualitative, involving frequency distributions, measures of central tendency and dispersion or quantitative, which demands the use of such statistical tools as simple percentage, frequency distribution, measures of central tendency, measures of dispersion. Quantitative analysis is used in summarising quantitative information generated by the research process.

3.3.2 Causal analysis

Causal analysis involves the use of more sophisticated statistical tools to draw inferences based on the research information. Such statistics as Chi-Squares, student t statistic, and least-squares estimators are some of the important statistical tools that can be used.

Causal analysis focuses on causal relationships between relevant research variables. The analytical results would help in isolating the causes of a given problem situation. Causal analysis also uses such statistical tools as correlation coefficient, test of goodness – of – fit, test of independence, and the like. It requires proficiency in statistical analysis. See Onwe (1998) and standard statistics texts for the application of the statistical tools mentioned here.

3.4 Tools for the Presentation and Analysis of Research Data

This section examines the relevant data presentation and analytical tools in business statistics. The tools include:

i. Frequency distribution tools
ii. Parametric tools
iii. Non-parametric tools

3.4.1 Frequency Distribution Tools

These tools basically involve the use of histograms and frequency polygons. In this section, you will be taken through the use of these tools.

Histogram and frequency polygon

A histogram is a pictorial representation of a frequency distribution of a given grouped data. The histogram is a set of bars or blocks constructed from the grouped data. The height of each bar is represented by the frequency of the corresponding observation, and the horizontal axis of each bar is represented by the class width.

The frequency polygon is a diagram obtained by connecting the midpoints of the bars or blocks to form the histogram of the given set of grouped data. For example, consider the frequency distribution of the monthly salary of 100 workers in a given company:

Table 6: Frequency Distribution of Monthly salaries of 100 workers

Monthly Salary (N'00s)	**No. of Workers** (f)
640 – 659	7
660 – 679	20
680 – 699	33
700 – 719	25
720 – 739	11
740 – 759	4
	$\sum f = 100$

The frequency distribution is represented by a two-dimensional graph, with the vertical axis labeled "frequencies (f)" and the horizontal axis labeled "Actual Class Limits". The actual class limits are obtained by subtracting a constant value, 0.5, from each lower class limit of the frequency distribution and adding same (0.5) to the corresponding upper class limits. This process is presented below.

Table 7: The Actual Frequency Distribution of Monthly Salaries of 100 Workers

Stated salary limits	Actual salary limits	No. of Workers
640-659	639.5-659.5	7
660-679	659.5-679.5	20
680-699	679.5-699.5	33
700-719	699.5-719.5	25
720-739	719.5-739.5	11
740-759	739.5-759.5	4

Table 7 is plotted in figure 8 to produce the histogram and polygon.

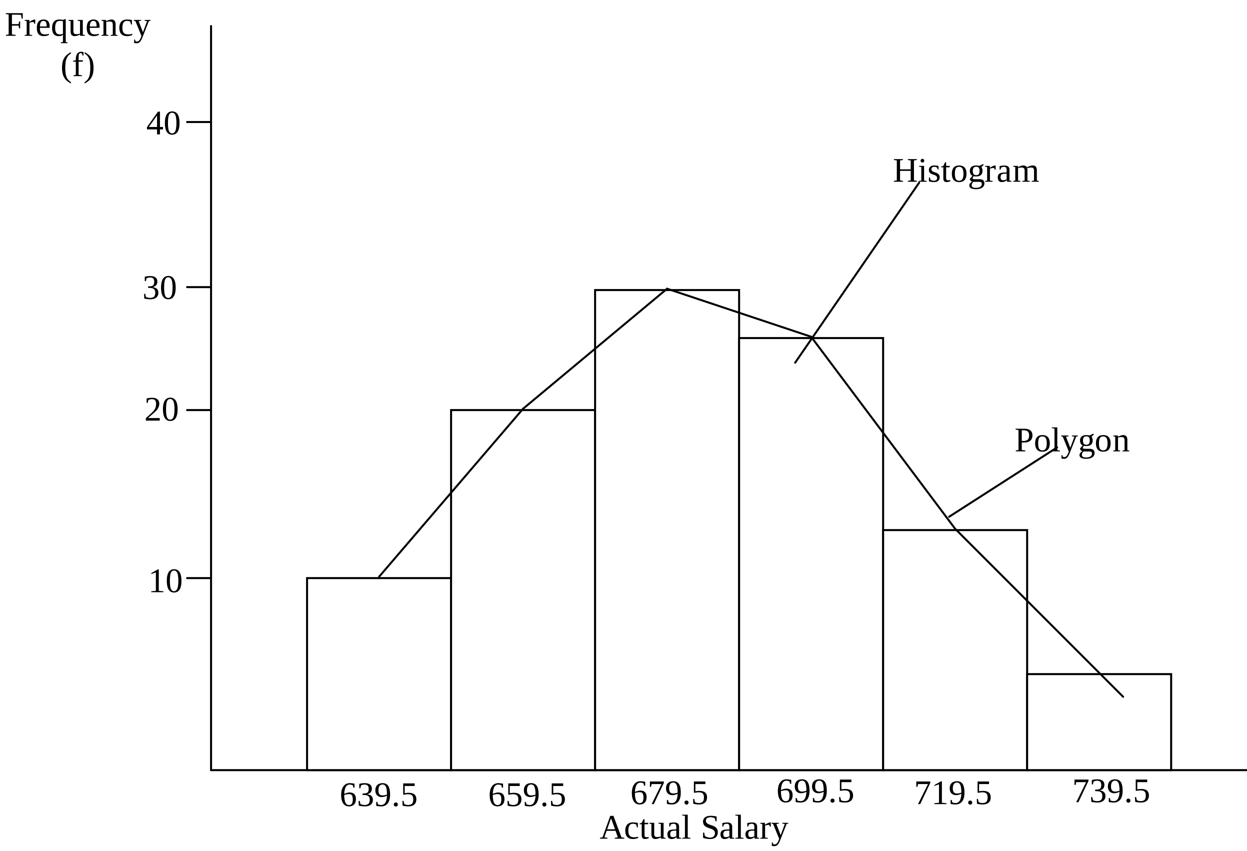


Fig. 8: The Graph of Histogram and Polygon

The Pie Chart

A pie chart shows the totality of the data being represented using a single circle (a "pie"). The circle is split into sectors, the size of each being drawn in proportion to the class frequency. For easy analysis, each sector is shaded differently.

For example, consider Table 8 below showing the non-managerial workforce at a given factory.

Table 8: Non-Managerial Workforce

Job Description	**Number Employed**
Labourers	21
Mechanics	38
Fitters	9
Clerks	12
Draughtsmen	4
Total	**84**

The pie chart for these data is as shown in Figure 9 below.

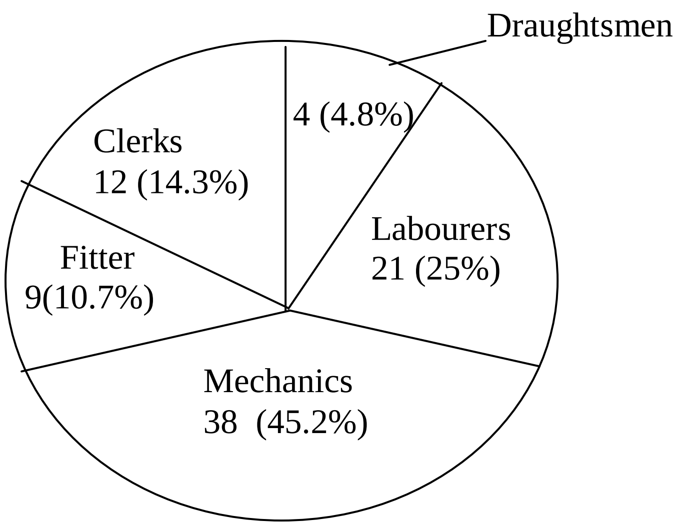


Fig. 9: Workforce Employed at a Factory

3.4.2 Parametric Tools

Our major interest here is on the tools used in testing parametric hypotheses. That is, hypotheses concerning population parameters, such as the population mean, μ, the population variance, δ^2 , and the population standard deviation, δ. For practical purposes, the major parametric tools are the student t-statistic and the Z-statistic. These were discussed in earlier. You will learn more on parametric tools in subsequent units.

3.4.3 Non-Parametric Tools

As presented in the last unit, the major parametric tool used in testing hypotheses is the chi-square ($\chi 2$)-statistic. Another is the Spearman rank correlation coefficient (r_s). The use of chi-square was discussed in detail earlier. You may also refer to it. The rank correlation coefficients will be discussed in subsequent units.

SELF-ASSESSMENT EXERCISE

Discuss the importance of field editing in the data analysis process.

4.0 CONCLUSION

This unit has worked on the necessary preliminaries in the processing, presentation, and analysis of research data. You were informed on the two basic types of data analysis: descriptive analysis and causal analysis. You also learned the basic tools of data analysis, including: frequency distribution tools, parametric tools; and, non-parametric tools.

5.0 SUMMARY

Before you present your research data for analysis and interpretation, it must be edited and coded. Two types of editing were discussed including: field editing; and, central editing. The editing process will enable you discover and correct possible errors in the data you collected. The errors can be either of the followings: numeric errors, errors of transposition, errors of inappropriate response, and errors of omission.

6.0 TUTOR-MARKED ASSIGNMENT

Obtain some real data on the distribution of sales revenue in a set of business organisations of your choice. Using the data, construct a frequency polygon.

7.0 REFERENCES/FURTHER READING

Nachimas, D. & Nachimas, C. (1976). *Research Methods in the Social Sciences.* New York: St Martin's Press.

Onwe, O. J. (2007). *Statistical Methods for Business and Economic Decisions: A Practical Approach*. Lagos: Samalice Press.

Nworgu, B. G. (1991) *Educational Research: Basic Issues and Methodology*. Ibadan: Wisdom Publishers Ltd.

Onwe, O. J. (2009) "Acceptability of Research Proposals in Social Sciences: Important Considerations in the West African Sub-Region." *Journal of Social Policy and Society,* vol. 4 (1).

Onwe, O. J. (1998). *Elements of Project and Dissertation Writing: A Guide to Effective Dissertation Report.* Lagos: Impressed Publishers.

Osuala, E. C. (1993). *Introduction to Research Methodology*. Onitsha: Africana-Fep Publishers Ltd.

MODULE 6 TOOLS OF DATA ANALYSIS AND DETERMINATION OF THE SAMPLE SIZE

UNIT 1 ADVANCED PARAMETRIC TOOLS OF DATA ANALYSIS: THE LINEAR AND RANK CORRELATION COEFFICIENTS

CONTENTS

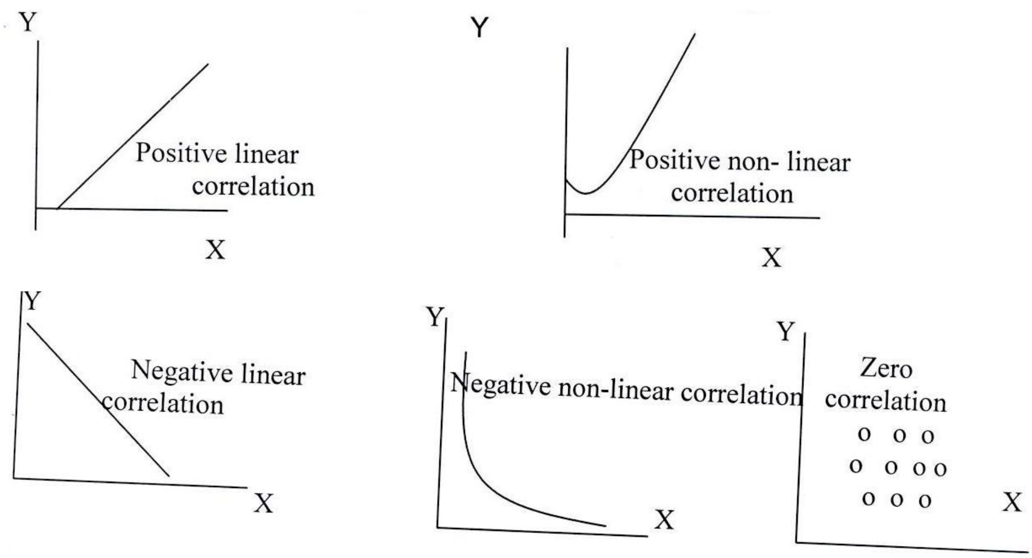
1.0 INTRODUCTION

The simplest methods of measuring relationships existing between economic variables are correlation analysis and regression analysis.

Correlation can be defined as the degree of relationship between two or more variables. The degree of relationship between two variables is called simple correlation. That connecting three_or more variables is called multiple correlation.

Correlation may be linear for scatter diagram on the values of two variables, (X and Y) are clustered near a straight line, or nonlinear, when all points on the scatter lie near a curve.

Two variables may have a positive correlation or a negative correlation, or they may be unrelated. These correlations are represented in the following diagrams:



This unit will expose you to two major coefficients often used in correlation analysis: the linear correlation coefficient and the rank correlation coefficient.

2.0 OBJECTIVES

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

- discuss the relationships between variables
- calculate correlation coefficients using statistical formulas
- interpret and apply correlation coefficient in data analysis.

3.0 MAIN CONTENT

3.1 The Linear Correlation Coefficients

You can determine the kind of correlation between two variables by direct observation of the scatter diagrams. In addition, the scatter diagram indicates the strength of the relationship between the two variables. If the points lie close to the line, the correlation is strong. A greater dispersion of points about the line implies weaker correlation.

Note that the inspection of a scatter diagram only gives a rough picture of the relationship between two variables. For a precise quantitative measurement of the degree of correlation between two variables, say X and Y, we use a parameter ρ referred to as the correlation coefficient. The sample estimate of this parameter is referred to as r.

The sample correlation coefficient is computed by:

$$r_{xy} = \frac{\sum(X_i - X)(Y_i - Y)}{\sqrt{\sum(X_i - X)^2}\sqrt{\sum(Y_i - Y)^2}}$$

$$= \frac{n\sum X_i Y_i - \sum(X_i)(\sum Y_i)}{\sqrt{n\sum X_i^2 - (\sum X_i)^2}\sqrt{n\sum Y_i^2 - (\sum Y_i)^2}}$$

Example: We want to determine the correlation between Price, X, and quantity supplied, Y, Given the following data:

Time Period (in days)	Quantity supply (Y_i) (in tons)	Unit Price (X_i) (in Naira)
1	10	2
2	20	4
3	50	6
4	40	8
5	50	10
6	60	12
7	80	14
8	90	16
9	90	18
10	120	20
n = 10	$\sum Y_i = 610$	$\sum X_i = 110$

$$X = \frac{\sum X_i}{n} = \frac{110}{10} = 11, \qquad Y = \frac{\sum Y_i}{10} = \frac{610}{10} = 61$$

It can be shown that the formula:

$$r = \frac{\sum(X_i - X)(Y_i - Y)}{\sqrt{\sum(X_i - X)^2}\sqrt{\sum(Y_i - Y)^2}}$$

$$= \frac{n\sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{n\sum X_i^2 - (\sum X_i)^2}\ \sqrt{n\sum Y_i^2 - (\sum Y_i)^2}}$$

Using the table:

X_i	Y_i	X_iY_i	X_i^2	Y_i^2
2	10	20	4	100
4	20	80	16	400
6	50	300	36	2,500
8	40	320	64	1,600
10	50	500	100	2,500
12	60	720	144	3,600
14	80	1120	196	6,400
16	90	1,440	256	8,100
18	90	1620	324	8100
20	120	2400	400	14400
$\sum X_i = 110$	$\sum Y_i = 610$	$\sum X_iY_i = 8520$	$\sum X_i^2 = 1540$	$\sum Y_i^2 = 47700$

$$r = \frac{n\sum X_iY_i - (\sum X_i)(\sum Y_i)}{\sqrt{n\sum X_i^2 - (\sum X_i)^2}\ \sqrt{n\sum Y_i^2 - (\sum Y_i)^2}}$$

$$= \frac{10(8520) - (110)(610)}{\sqrt{10(1540) - (110)^2}\ \sqrt{10(47700) - (610)^2}}$$

$$= \frac{85200 - 67100}{\sqrt{15400 - 12100}\ \sqrt{477000 - 372100}}$$

$$= \frac{18100}{\sqrt{3300}\ \sqrt{104900}} = \frac{18100}{(57.45)(323.88)}$$

$$= \frac{18100}{18606.91} = 0.972 \text{ approx.}$$

3.2 The Rank Correlation Coefficient

Rank correlation coefficient is used for qualitative variables, whereby the variables_cannot be measured numerically. It can be used in testing non-parametric hypotheses. Examples of such variables include profession, education, preferences for a particular brand of commodity and the like.

In such cases, we rank the observations in a specific sequence, such as in order of_size, importance, etc., using numbers 1,2,…, n. We therefore, assign ranks to the data and measure the relationships between their ranks instead of their actual numerical values.

If two variables, X and Y are ranked, the rank correlation coefficient is computed by:

$$r_s = 1 - \frac{6\sum d^2}{n^3 - n}$$

where $d^2 = [R(X)-R(Y)]^2$ = squared difference between ranks of corresponding pairs of X and Y.

R(X) = individual ranks assigned to X values
R(Y) = individual ranks assigned to Y values
n = number of pairs of observations

There are two important points to note about the rank correlation coefficient:

1. It does not matter whether the ranking of the observations is in ascending or descending order.
2. If two or more observations have the same value, you will assign the **mean rank** to each observation.

Examples:

1. Consider the following table showing how 10 students were ranked according to their performance in their class work and their final examinations. Let us find out whether there is a relationship between the accomplishments of the students during the year and their performance in their examinations.

Student	Ranking Based on Class Work	Ranking Based on Exam Marks	d	d^2
A	2	1	1	1
B	5	6	-1	1
C	6	4	2	4
D	1	2	-1	1
E	4	3	1	1
F	10	7	3	9
G	7	8	-1	1
H	9	10	-1	1
I	3	5	-2	4
J	8	9	-1	1
				$\sum d^2 = 24$

The correlation coefficient in this example is:

$$r_s = 1 - \frac{6\sum d^2}{n^3 - n} = 1 - \frac{6(24)}{10^3 - 10} = 1 - 0.145 = 0.855$$

The value of $r_s = 0.855$ indicates a close relationship between class work and examination performance, so that students with good records throughout the year do well in their examination and vice versa.

2. In a recent research project, a student attempted to find out if the demand for one specific product is affected by expenditures on television advertisements.

The following data were obtained from the reference firm:

Year	Qd (in thousands) (Q)	Expenditure on TV Adverts (X) (N'millions)	R(X)	R(Q)	d R(x)-R(Q)	d^2 $\{R(x) – R(Q)\}^2$
1980	35.00	0.05	3.0	3.5	-0.50	0.025
1981	37.50	0.09	7.0	5.0	2.00	4.00
1982	39.20	0.04	1.5	7.0	-5.50	30.25
1983	22.00	0.04	1.5	1.0	0.50	0.25
1984	33.00	0.10	8.5	2.0	6.50	42.25
1985	41.60	0.10	8.5	8.0	0.50	0.25
1986	44.80	0.11	10.	9.0	1.00	1.00
1987	35.00	0.08	6.0	3.5	2.50	6.25
1988	38.00	0.06	4.0	6.0	-2.00	4.00
1989	51.80	0.12	11.0	10.0	1.00	1.00
1990	62.05	0.14	13.0	12.0	1.00	1.00
1991	63.00	0.13	12.0	13.0	-1.00	1.00
1992	58.30	0.07	5.0	11.0	-6.00	36.00
1993	70.00	0.20	14.0	15.0	-1.00	1.00
1994	68.08	0.21	15.0	14.0	-1.00	1.00
1995	72.00	0.27	16.0	16.0	0.00	0.00
						$\sum d^2=129.50$

The rank correlation coefficient can be computed as follows:

$$r_s = 1 – \frac{6\sum d^2}{n^3-n} = 1 – \frac{6(129.50)}{16^3 – 16}$$

$$= 1 – \frac{777}{4080} = 0.81$$

The value, $r_s = 0.81$ indicates a strong positive correlation between the quantity demanded of the specific product (Q) and expenditure on TV adverts (X).

SELF-ASSESSMENT EXERCISE

Identify two quantitative variables of interest, obtain about 10 values for each variable, plot these values on a two-dimensional graph, and describe the observed relationship between the two variables from the plotted graph.

4.0 CONCLUSION

This unit has exposed you to the correlation theory. It defined correlation as the degree of relationship between two variables. It also informs you that when you connect three or more variables you get what is referred to as multiple correlations. Two major useful correlation concepts were discussed including the simple linear correlation and the rank correlation. The formulas for computing the correlation coefficients were discussed at large, with relevant examples.

5.0 SUMMARY

The simplest methods of measuring relationships existing between economic variables involve correlation analysis and regression analysis.

Correlation can be defined as the degree of relationship between two or more variables. The degree of relationship between two variables is called simple correlation. Those connecting three_or more variables are called multiple correlations.

Correlation may be linear for scatter diagram on the values of two variables, (X and Y) are clustered near a straight line, or nonlinear, when all points on the scatter lie near a curve.

The type of correlation between two variables can be determined by direct observation of the scatter diagrams. The scatter diagram indicates the strength of the relationship between the two variables. If the points lie close to the line, the correlation is strong. A greater dispersion of points about the line implies weaker correlation.

Rank correlation coefficient is used for qualitative variables, whereby the variables_cannot be measured numerically. It can be used in testing non-parametric hypotheses. Examples of such variables include profession, education, preferences for a particular brand of commodity and the like.

6.0 TUTOR-MARKED ASSIGNMENT

The following data give the actual sales of a company in each of the eight states of a hypothetical country, together with the forecast of sales by two different methods:

State	Actual Sales (N'million)	Forecast 1	Forecast 2
A	15	13	16
B	19	25	19
C	30	23	26
D	12	26	14
E	58	48	65
F	10	15	19
G	23	28	27
H	17	10	22

(a) Calculate the rank correlation coefficient between:
   (i) Actual sales and forecast 1
   (ii) Actual sales and forecast 2

(b) Which forecasting method would you recommend next year and why?

7.0 REFERENCES/FURTHER READING

Onwe, O. J. (2007). *Statistical Methods for Business and Economic Decisions: A Practical Approach*. Lagos: Samalice Press.

Onwe, O. J. (2012). *Fundamentals of Managerial Statistics and Econometric Analysis: Some Practical Approaches.* Lagos: Impressed Publishers.

Chinelo Obeleagu-Nzelibe (ed.). (1995). *Business Statistics: Theory and Applications.* Enugu: Optimal Publishers.

UNIT 2 ADVANCED PARAMETRIC TOOLS OF ANALYSIS: THE PARTIAL CORRELATON COEFFICIENTS

CONTENTS

1.0 INTRODUCTION

A partial correlation coefficient measures the relationship between any two variables, keeping other variables constant. It is a very useful coefficient in data analysis as researchers are often interested in knowing specific impacts of one variable in a pull of different variables determining specific behaviours. This unit will help you understand the computation, use, and basis for the concept of partial correlation.

2.0 OBJECTIVES

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

- explain correlation coefficients
- differentiate between simple linear correlation and partial correlation
- discuss how to compute correlation coefficients
- apply the concept of partial correlation in data analysis

3.0 MAIN CONTENT

3.1 Computation of the Partial Correlation Coefficients

Assume a multiple relationship between three variables, X_1, X_2, and X_3. To measure the true correlation between X_1 and X_2, we find the partial correlation coefficient between X_1 and X_2, keeping X_3 constant.

The partial correlation coefficient is determined in terms of the simple correlation coefficients among the various variables involved in a multiple relationship. For the three variables, X_1, X_2 and X_3, three simple correlation coefficients are involved as follows:

(1) r_{12} = correlation coefficient between X_1 and X_2
(2) r_{13} = correlation coefficient between X_1 and X_3
(3) r_{23} = correlation coefficient between X_2 and X_3.

There are also two partial correlation coefficients involved:

(1) $r_{12.3}$ = partial correlation coefficient between X_1 and X_2, keeping X_3 constant.

This is given by:

$$r_{12.3} = \frac{r_{12} - (r_{13})(r_{23})}{\sqrt{[1-(r_{13})^2][1-(r_{23})^2]}}$$

(2) $r_{13.2}$ = partial correlation coefficient between X_1 and X_3, keeping X_2 constant:

$$r_{13.2} = \frac{r_{13}-(r_{12})(r_{23})}{\sqrt{[1-(r_{12})^2][1-(r_{23})^2]}}$$

The formula for partial correlation coefficient can be extended to relationships involving any number of explanatory variables.

Examples

An automobile dealer observes the following results on his business variables for 6 months:

	Month					
Variable	1	2	3	4	5	6
X_1	10	10	20	30	40	40
X_2	0	1	2	2	3	4
X_3	1	0	2	3	3	3

where,

X_1 = number of new cars sold per month
X_2 = number of 10 minute local TV Spot Advert during the month
X_3 = number of full-page newspaper advert during the month.

(a) Compute the simple correlation coefficient:
  i) Between X_1 and X_2, (r_{12})
  ii) Between X_1 and X_3, (r_{13})
  iii) Between X_2 and X_3, (r_{23})

(b) Compute the Partial Correlation Coefficient:
  i) Between X_1 and X_2, keeping X_3 constant, ($r_{12.3}$)
  ii) Between X_1 and X_3, keeping constant, ($r_{13.2}$)

Solutions

A (i)
$$r_{12} = \frac{n\sum X_1X_2 - \sum X_1\sum X_2}{\sqrt{n\sum X^2_1 - (\sum X_1)^2}\ \sqrt{n\sum X^2_2 - (\sum X_2)^2}}$$

	X_1	X_2	X_1^2	X_2^2	X_1X_2
	10	0	100	0	0
	10	1	100	1	10
	20	2	400	4	40
	30	2	900	4	60
	40	3	1600	9	120
	40	4	1600	16	160
Total	150	12	4700	34	390

$n\sum X_1X_2 = 6(390) = 2340$
$\sum X_1\sum X_2 = 150(12) = 1800$
$n\sum X_1^2 = 6(4700) = 28,200$
$n\sum X_2^2 = 6(34) = 204$
$(\sum X_1)^2 = (150)^2 = 22500$
$(\sum X_2)^2 = (12)^2 = 144$

$$r_{12} = \frac{2340 - 1800}{\sqrt{28200-22500}\ \sqrt{204-144}}$$

$= 540/(75.4983)(7.7460)$

$= 540/584.81 = 0.923$

This implies a very close relationship between the number of cars sold per month and the number of 10-minute local TV Sport adverts during the month.

a (ii) $r_{13} = \dfrac{n\Sigma X_1X_3 - \Sigma X_1\Sigma X_3}{\sqrt{n\Sigma X_1^2-(\Sigma X_1)^2}\sqrt{n\Sigma X_3^2-(\Sigma X_3)^2}}$

Using the table below:

X_1	X_3	X_1^2	X_3^2	X_1X_3
10	1	100	1	10
10	0	100	0	0
20	2	400	4	40
30	3	900	9	90
40	3	1600	9	120
40	3	1600	9	120
Total =150	**12**	**4700**	**32**	**380**

$n\Sigma X_1X_3 = 6(380) = 2280$

$\Sigma X_1\Sigma X_3 = 150(12) = 1800$

$n\Sigma X_1^2 = 6(4700) = 28200$

$n\Sigma X_3^2 = 6(32) = 192$

$(\Sigma X_1)^2 = (150)^2 = 22500$

$(\Sigma X_3)^2 = (12)^2 = 144$

$r_{13} = \dfrac{2280 - 1800}{\sqrt{28200 - 22500}\sqrt{192 - 144}}$

$= \dfrac{480}{(75.4983)(6.9282)}$

$= 480/523.0673 = 0.918$

Again, the coefficient implies a strong relationship between number of new cars sold per month and number of full page newspaper advert during the month.

a (iii) $r_{23} = \dfrac{n\Sigma X_2X_3 - \Sigma X_2\Sigma X_3}{\sqrt{n\Sigma X_2^2 - (\Sigma X_2)^2}\sqrt{n\Sigma X_3^2 - (\Sigma X_3)^2}}$

The expressions in the formula can be calculated using the following table:

X_2	X_3	X_2^2	X_3^2	X_2X_3
0	1	0	1	0
1	0	1	0	0
2	2	4	4	4
2	3	4	9	6
3	3	9	9	9
4	3	16	9	12
Total = 12	**12**	**34**	**32**	**31**

$n\Sigma X_2X_3 = 6(31) = 186$

$\Sigma X_2 \Sigma X_3 = 12(12) = 144$
$n\Sigma X_2^2 = 6(34) = 204$
$(\Sigma X_2)^2 = (12)^2 = 144$
$n\Sigma X_3^2 = 6(32) = 192$
$(\Sigma X_3)^2 = (12)^2 = \; = 144$

$$r_{23} = \frac{186 - 144}{\sqrt{(204-144)}\sqrt{(192-144)}}$$

$$= 42/(7.746)(6.9282)$$
$$= 42/53 = 0.783$$

b. (i) Using the figures obtained in a(i), a(ii), and a(iii) above, we compute the partial correlation coefficients, $r_{12.3}$ as follows:

$$r_{12.3} = \frac{r_{12} - (r_{13})(r_{23})}{\sqrt{[1 - (r_{13})^2][1 - (r_{23})^2]}}$$

$$= \frac{0.923 - (0.918)(0.783)}{\sqrt{[1 - (0.918)^2][1 - (0.783)^2]}}$$

$$= \frac{0.923 - 0.7188}{\sqrt{(0.1573)(0.3869)}}$$

$$= \frac{0.2042}{\sqrt{0.6082}}$$

$$= 0.2042/0.2467 = 0.8277$$

It follows that the partial correlation between the number of new cars sold per month, X_1, and the number of 10-minute local TV spot advert during the month, X_2, keeping the number of full-page advert, X_3, constant is high.

b. (ii) The partial correlation between the number of new cars sold per month and the number of full-page newspaper advert during the month, keeping the number of 10-minute advert during the month, $r_{13.2}$, can be similarly obtained as follows:

$$r_{13.2} = \frac{r_{13} - (r_{12})(r_{23})}{\sqrt{[1 - (r_{12})^2][1 - (r_{23})^2]}}$$

$$= \frac{0.918 - (0.923)(0.783)}{\sqrt{[1 - (0.923)2][1 - (0.783)2]}}$$

$$= \frac{0.1953}{\sqrt{(0.1489)(0.3869)}}$$

$$= 0.1953/0.2400 = 0.8138$$

The computed coefficient indicates a high partial correlation between the number of new cars sold and the number of full-page advert during the month, keeping the number of 10-minute spot advert constant.

3.2 Limitations of the Theory of Linear Correlations

The limitations of linear correlations as a technique for the study of economic relations are as follows:

1. The formula for correlation coefficient applies only to linear relationships between variables.
2. That correlation coefficient is a measure of co-variability of variables does not imply any functional relationship between the variables concerned.

SELF-ASSESSMENT EXERCISE

With appropriate example(s), discuss the basic differences between simple linear correlation and partial correlations.

4.0 CONCLUSION

You have just looked at another dimension to the concept of correlations. Partial correlation deals with correlation between two variables in a pull of multiple variables in data analysis. With simple examples the unit has illustrated the necessary formulas for computation of partial correlations.

The unit also noted two major limitations of the theory of linear correlations in general. The limitations are: (i) the observation that the formula for the computation of correlation coefficient applies only to linear relationships between variables; and, (ii) the observation that a correlation coefficient is a measure of co-variability of variables does not imply any functional relationship between the variables concerned.

5.0 SUMMARY

A partial correlation coefficient measures the relationship between any two variables, keeping other variables constant. It is determined in terms of the simple correlation coefficients among the various variables involved in a multiple relationship.

For the three variables, X_1, X_2 and X_3, you can obtain three simple correlation coefficients are:

(1) r_{12} = correlation coefficient between X_1 and X_2
(2) r_{13} = correlation coefficient between X_1 and X_3
(3) r_{23} = correlation coefficient between X_2 and X_3.

This will give rise to two partial correlation coefficients:

(1) $r_{12.3}$ = partial correlation coefficient between X_1 and X_2, keeping X_3 constant.

(2) $r_{13.2}$ = partial correlation coefficient between X_1 and X_3, keeping X_2 constant.

There are two basic limitations of linear correlations as a technique for the study of economic relations, these are:

1. The formula for correlation coefficient applies only to linear relationships between variables.
2. That correlation coefficient is a measure of co-variability of variables does not imply any functional relationship between the variables concerned.

6.0 TUTOR-MARKED ASSIGNMENT

Using the following variables:

X_1 = the quantity of commodity A sold per month

X_2 = number of sales representatives engaged per month
X_3 = number of full-page newspaper advert during the month.

Make hypothetical assignment of values to these variables for 10 months as guided by the following:

	Month									
Variable	1	2	3	4	5	6	7	8	9	10
X_1										
X_2										
X_3										

Compute the partial correlation coefficient:

i) Between X_1 and X_2, keeping X_3 constant, ($r_{12.3}$)
ii) Between X_1 and X_3, keeping constant, ($r_{13.2}$)

7.0 REFERENCES/FURTHER READINGS

Chinelo Obeleagu-Nzelibe (ed.). (1995). *Business Statistics: Theory and Applications.* Enugu: Optimal Publishers.

Nachimias, D. & Nachimias, C. (1976). *Research Methods in The Social Sciences.* New York: Martin's Press.

2. Onwe, O. J. (2007). *Statistical Methods for Business and Economic Decisions: A Practical Approach*. Lagos: Samalice Press.

Onwe, O. J. (2012). *Fundamentals of Managerial Statistics and Econometric Analysis: Some Practical Approaches.* Lagos: Impressed Publishers.

UNIT 3 ADVANCED PARAMETRIC TOOLS OF ANALYSIS: THE SIMPLE LINEAR REGRESSIONS

CONTENTS

1.0 INTRODUCTION

Simple regression analysis has been variously applied in social science research. It is one of the important scientific methods used in data analysis, mostly focusing on causal relationships. This unit will summarise the basic principles of regression analysis and acquaint you with the practical aspects of the use of simple linear regressions.

2.0 OBJECTIVES

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

- apply the concepts of linear regressions
- carry out statistical analysis
- enumerate the basic differences between correlations and regressions
- describe the functional forms of regression models

- highlight the necessary formulas and practical applications of regression analysis.

3.0 MAIN CONTENT

The features of regression analysis will be discussed in terms of six major headings:

- Origin of “regression”
- Modern interpretation of “regression”
- Statistical vs. Deterministic Relationships
- Regression vs. Causation
- Regression vs. Correlation
- Terminology and Notation

3.1 Origin of the Term Regression

The term ‘regression’ originated from an author by the name Francis Galton in 1886. Galton found that, although there was a tendency for tall parents to have tall children and short parents to have short children, the average height of children born of parents of a given height tended to move or “regress” towards the average height in the population as a whole. In other words, the height of the children of unusually tall or unusually short parents tends to move toward the average height of the population. Galton referred to this as the **law of universal regression**.

3.2 Modern Interpretation of Regression

According to the modern interpretation, regression analysis is concerned with the study of the dependency of one variable on other variables, with a view to estimating and/or predicting the population average value of the dependent variable in terms of known or fixed values of the independent variable.

3.3 Statistical Vs Deterministic Relationships

Statistical relationships deal with random or stochastic variables, that is, variables that are probabilistic in nature or have probability distributions. In functional or deterministic relationships, we deal with non-random or non-stochastic variables.

The dependence of crop yield on temperature, or rainfall, for instance, is statistical in nature because the explanatory variables (that is, temperature and rainfall in this case), though certainly important, will not enable the agronomist predict exactly the crop yield due to the uncertain nature of the explanatory variables.

In deterministic relationships, we deal with relationships of the type exhibited by Newton's law of gravity, for example. Recall that the Newton's law states that "Every particle in the universe attracts every other particle with a force directly proportional to the product of their masses and inversely proportional to the square of the distance between them." Symbolically, $F = k(m_1m_2/r^2)$, where F = force; m_1 and m_2 are the masses of the two particles; r = distance; and, k = constant of proportionality.

3.4 Regression Vs Causation

Though regression analysis deals with the dependence of one variable on other variables, it does not necessarily imply causation. In the crop yield example above, there is no statistical reason to assume that rainfall does not depend on crop yield. The fact that we treat crop yield as dependent on rainfall is as a result of non-statistical considerations: Common sense would suggest that the relationship cannot be reversed, for we cannot control rainfall by varying crop yield.

3.5 Regression Vs Correlation

The primary objective of correlation is to measure the strength or degree of linear relationship between two variables. The correlation coefficient measures the strength of linear association. For example, we may be interested in finding the correlations coefficient between smoking and lung cancer, between scores on statistics and mathematics examinations, etc. In regression analysis, we are not primarily interested in such measure. Instead, we try to estimate or predict the average value of one variable on the basis of the fixed values of other variables. We may want to know whether we can predict the average score on a statistics examination by knowing a student's score on a mathematics examination.

3.6 Terminology and Notation

An understanding of the necessary terminologies and notations in regressions is necessary for the understanding of regression analysis. The terms, dependent variable and explanatory variable are common in econometrics. Most econometric texts use the dependent-variable-explanatory-variable terminology. A representative list of such terms is summarised as follows:

Dependent Variable	Explanatory Variable
↑↓	↑↓
Explained Variable	Independent Variable
↑↓	↑↓
Predictand	Predictor
↑↓	↑↓
Regressand	Regressor
↑↓	↑↓
Response	Stimulus or Control Variable
↑↓	↑↓
Endogenous	Exogenous

If the analysis involves a dependent variable and one independent or explanatory variable, such as consumption as the dependent variable and income as the independent variable, we say that such analysis is simple or two-variable regression analysis. If the explanatory variables are two or more, we say that the analysis is multiple regression analysis.

The term random is a synonym for the term stochastic. A random or stochastic variable, as you may recall, is a variable that can take on any set of value, positive or negative, with a given probability.

3.7 Functional Forms of Regression Models

Apart from general linear models of the form, Y = a + bY, econometric researchers use the following regression models:

- The Log-linear model
- Semi-log models
- Reciprocal models

We examine the special features of these models.

3.7.1 The Log-Linear Model (Used in Measuring Elasticity)

Consider the following **exponential regression** model:

$$Y_i = \beta_1 X_i^{\beta 2} e^{ui} \tag{3.1}$$

Which can be expressed alternatively as:

$$\ln Y_i = \ln\beta_1 + \beta_2 \ln X_i + ui \tag{3.2}$$

where ln = natural log.

If this expression is written as:

$$\ln Y_i = \alpha + \beta_2 \ln X_i + ui, \qquad (3.3)$$

Where $\alpha = \ln \beta_1$, this model will be said to be linear in the parameters α and β_2, linear in the logarithms of the variables Y and X, and can be estimated by ordinary least-square (OLS) regression.

A unique feature of the log-linear model is that the slope coefficient, $\beta 2$ measures the elasticity of Y with respect to X, that is, the percentage change in Y for a given percentage change in X.

3.7.2 The Semi-Log Models

Economists, governments, and businesspeople alike are often interested in finding the rate of growth of certain economic variables, such as GNP, money supply, population, employment, trade deficit, productivity, and the like. Econometric estimate of these kinds of growth rate can be done using semi-log models exemplified as follows:

Recall the following well-known compound rate formula from the introductory courses in money, banking, and finance:

$$Y_t = Y_o(1 + r)^t \qquad (3.4)$$

where r = compound rate of growth of Y.

Taking natural log on both sides of the equation (3.4), we get:

$$\ln Y_t = \ln Y_o + t \ln(1 + r) \qquad (3.5)$$

Letting $\beta_1 = \ln Y_o$

$\beta_2 = \ln(1 + r)$, equation (3.5) can be written as:

$$\ln Y_t = \beta_1 + \beta_2 t \qquad (3.6)$$

Adding disturbance term to equation (3.6), we get:

$$\ln Y_t = \beta_1 + \beta_2 t + u_t \qquad (3.7)$$

The model represented by equation (3.7) is like any other linear model since the parameters $\beta 1$ and $\beta 2$ are linear. The difference is that the dependent variable (regresand), Yt, is the logarithm of Y, and the independent variable (regressor), t, is time. Models like that of equation (3.7) are referred to as semi-log models.

Researchers, instead of estimating model (3.7), sometimes estimate the following model:

$$Y_t = \beta_1 + \beta_2 t + u_t \qquad (3.8)$$

Models represented by equation (3.8) are called linear trend models. Here, the time variable t is known as the trend variable.

3.7.3 The Reciprocal Model

Reciprocal models are of the following form:

$$Y_i = \beta_1 + \beta_2(1/X_i) + u_i \qquad (3.9)$$

Observe that, though the model represented by equation (3.9) is nonlinear in the variable X since it enters inversely or reciprocally, it is linear in β1 and β2 and is therefore a linear regression model.

3.8 The Least Squares and the Normal Equation of Ordinary Least-Squares (OLS) Equation

Assume two variables, X and Y. Denote the sample observations by:

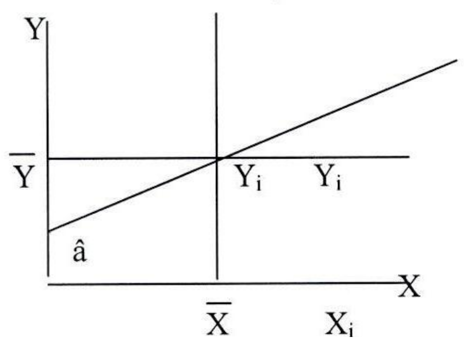
for X: $X_1, X_2, X_3, \ldots, X_n$

for Y: $Y_1, Y_2, Y_3, \ldots, Y_n$

Their respective means are: $\bar{X} = \frac{\sum X}{n}$; $\bar{Y} = \frac{\sum Y}{n}$

Representing these mean on a scatter diagram, we can plot the estimated line,

$Y = \hat{a} + bX$ as follows:



Taking a point P with coordinate points (X_i, Y_i) in the scatter diagram, we define the vertical difference between P and the estimated line by:

$$ei = Y_i - Y_i \qquad \text{(residuals)}$$

These residuals or deviations from the estimated line will be positive or negative as the actual point lies above or below the line. If they are squared and summed, the resultant quantity must be positive. Different pairs of values for a and b will give different lines and hence different values for the sum of the squared residuals about the line.

Thus, $\sum e_i^2 = f(\hat{a},b)$

The principle of least squares is that the â,b values should be chosen so as to minimise $\sum e_i^2$. The first order condition for a minimum requires that:

$$\frac{\partial}{\partial a}(\sum e_i^2) = 0, \text{ and,}$$

$$\frac{\partial}{\partial b}(\sum e_i^2) = 0$$

By substitution,

$$\begin{aligned}\sum e_i^2 &= \sum(Y_i - Y_i)^2 \\ &= \sum(Y_i - (a+bX_i))^2 \\ &= \sum(Y_i - a - bX_i)^2\end{aligned}$$

Thus, $\sum e_i^2 = \sum(Y_i - \hat{a} - bX_i)^2$

$$\frac{\partial}{\partial \hat{a}}(\sum e_i^2) = -1[2\sum(Y_i - \hat{a} - bXi)]$$

$$= -2\sum(Yi - \hat{a} - bX_i) = 0 \qquad (1)$$

$$\frac{\partial}{\partial b}(\sum e_i^2) = -X_i [2\sum(Y_i - \hat{a} - bXi)]$$

$$= -2\sum Y_i(Y_i - \hat{a} - bX_i) = 0 \qquad (2)$$

Simplifying equations (1) and (2), we get the standard form of the normal equations for a straight line as follows:

Simplifying equation (1):

$$-2\sum(Y_i - \hat{a} - bX_i) = 0$$

$$\sum(Y_i - \hat{a} - bX_i) = 0$$

$$\sum Y_i - n\hat{a} - b\sum X_i = 0$$
$$\sum Y_i = n\hat{a} + b\sum X_i \qquad (3)$$

Simplifying equation (2):

$$-2\sum X_i(Y_i - \hat{a} - bX_i) = 0$$
$$\sum X_i(Y_i - \hat{a} - bX_i) = 0$$
$$\sum (X_iY_i - \hat{a}\, X_i - bX_i^2) = 0$$
$$\sum X_iY_i - \hat{a}\sum X_i - b\sum X_i = 0$$
$$\sum X_iY_i = a\sum X_i + b\sum X_i^2 \qquad (4)$$

Equation (3) and (4) are referred to as the normal equation of OLS. Inserting the indicated values from the sample observations into equations (3) and (4) we get two simultaneous equations which can be solved for a and b.

Alternatively, if we divide equations (3) by n, we get:

$$\frac{\sum Y_i}{n} = \frac{n\hat{a}}{n} + \frac{b\sum X_i}{n}$$
$$Y = \hat{a} + bX \qquad (5)$$

This implies that the least squares estimates are such that the estimated line passes through the point of means (X,Y).

Subtracting equation (5) from the estimated line,

$Y = \hat{a} + bX$, we get:

$$Y = \hat{a} + bX$$
$$-(Y=\hat{a} + bX)$$
$$Y - Y = bX - bX$$
$$Y - Y = b(X - X) \qquad (6)$$

Using lower case letters to denote deviations from the mean, we get:

$x_i = X_i - X_i$; $y_I = Y - Y$; $y_i = Y - Y$

The equation of the least-squares line can now be written as:

$y = bx$ (from equation 6)

The deviation or residual becomes:

$ei = y_i - y_i = y_i - bx_i$

and the sum of squares residual becomes:

$$\Sigma e_i^2 = \sum(y_i - bx_i)^2 \qquad (7)$$

With the first order condition, we minimise equation (7) with respect to b to get:

$$\frac{\partial}{\partial b}(\sum e_i)^2 = \frac{\partial}{\partial b}\sum(y_i - bx._i)^2$$

$$-2u_i\sum(y_i - bx_i) = 0$$

$$x_i\sum(y_i - bx_i) = 0$$

$$\sum x_i y_i - b\sum x_i^2 = 0$$

$$b\sum x_i^2 = \sum x._i y_i$$

$$b = \frac{\sum x_i y_i}{\sum x_i^2} \qquad (8)$$

Since, $x._i = X_i - X$

$y_i = Y_i - Y$

it follows that equation (8), becomes:

$$b = \frac{\sum(X_i - X)(Y_i - Y)}{\sum(X_i - X)^2} \qquad (9)$$

Using equation (5), we get:

$$a = Y - bX$$

(10)

Equation (9) can be simplified to obtain:

$$b = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2}$$

Example

Estimate the tax linear equation for the following tax and income schedule:

Year	Income (X) (N'million)	Tax Revenue(Y) (N'million)
1964	26,934	4245
1965	28,729	5029
1966	30,171	5518
1967	31,781	5960
1968	33,450	6724

$Y = \hat{a} + bX$ = Tax Functions

Solution

By definition, $b = \dfrac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2}$

$$\hat{a} = Y - bX$$

X_i(N100s of million)	Y_i(N100s of million)	X_iY_i	X_i^2
269.34	42.45	11433.483	72544.035
287.29	50.29	14447.814	82535.544
301.71	55.18	16648.357	91028.924
317.81	59.60	18941.476	101003.19
334.50	67.24	22491.78	111890.25
$\sum Xi = 1510.65$	$\sum Y_i = 274.76$	$\sum X_iY_i = 83962.91$	$\sum X_i^2 = 459001.93$

Mean of X = <u>302.13</u> mean of Y = <u>54.95</u>

$$b = \frac{5(83962) - (1510.65)(274.76)}{5(459001.93) - (1510.65)^2} = \frac{419810 - 415066.19}{2295009.6 - 2282063.4}$$

$$= \frac{4743.81}{12946.5} = \underline{0.37} \text{ approx.}$$

$$\hat{a} = Y - bX = 54.95 - 0.37(302.13) = 54.95 - 111.79$$
$$= -56.84$$

The estimated line equation for the tax function is:

$$Y = -56.84 + 0.37X$$

SELF-ASSESSMENT EXERCISE

Discuss briefly, the differences between:

(a) Statistical and deterministic relationships
(b) Regression and causation
(c) Regression and correlation

4.0 CONCLUSION

This unit has presented the practical principles of regression analysis. The discussions also featured the following:

- origin of “regression”
- modern interpretation of “regression”
- statistical vs. deterministic relationships
- regression vs. causation
- regression vs. correlation

- terminology and notation

The functional forms of regression models were also discussed at length.

5.0 SUMMARY

The term 'regression' originated from an author by the name Francis Galton in 1886. Galton found that, although there was a tendency for tall parents to have tall children and short According to the modern interpretation, regression analysis is concerned with the study of the dependency of one variable on other variables, with a view to estimating and/or predicting the population average value of the dependent variable in terms of known or fixed values of the independent variable.

6.0 TUTOR-MARKED ASSIGNMENT

The data given below represent the annual gross income (in N' millions) obtained by a company over the periods t = 1 to t = 10:

Annual Gross Revenues

Year	Gross Income (N' million)
1	13.0
2	14.1
3	15.7
4	17.0
5	18.4
6	20.9
7	23.5
8	26.2
9	29.0
10	32.8

Using the Ordinary least squares principles estimate the relationship between the gross income (Y) and the time period (t), using the above data.

7.0 REFERENCES/FURTHER READING

Chinelo Obeleagu-Nzelibe (ed.). (1995). *Business Statistics: Theory and Applications.* Enugu: Optimal Publishers.

Nachimias, D. & Nachimias, C. (1976). *Research Methods in The Social Sciences*. New York: Martin's Press.

Onwe, O. J. (2007). *Statistical Methods for Business and Economic Decisions: A Practical Approach.* Lagos: Samalice Press.

Onwe, O. J. (2012). *Fundamentals of Managerial Statistics and Econometric Analysis: Some Practical Approaches.* Lagos: Impressed Publishers.

UNIT 4 POPULATION, SAMPLING, AND DETERMINATION OF THE SAMPLE SIZE

CONTENTS

1.0 INTRODUCTION

The social science research is based on the sampling of behaviours from a given population. The attempt is to make good representative sample of the population. In this unit, you will be led through the process of sampling with specific emphasis on sampling techniques. We begin by looking at the differences between a population and a sample.

2.0 OBJECTIVES

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

- distinguish between a population and a sample
- apply the different techniques of sampling
- determine the appropriate sample size for your chosen research activity.

3.0 MAIN CONTENT

3.1 Population and Sample

The term population depicts the totality of all behaviours in a given universe. It is synonymous to the universal set of all behaviours.

A sample is a subset of a population. Statisticians use sample behaviours to draw inferences about the population behaviour.

For a quick comparison, consider figure 10 below.

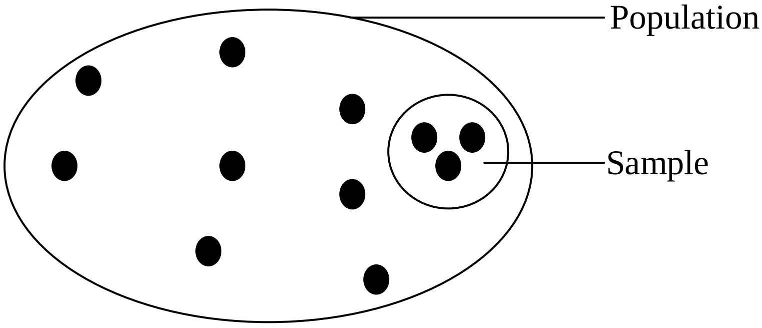


Fig. 10: **Pictorial Representation of the Difference between Population and Sample**

3.2 Sampling Techniques

There are two basic classifications of sampling techniques:

i. Probabilistic Sampling Techniques
ii. Non-Probabilistic Sampling Techniques

3.2.1 Probabilistic Sampling Techniques

1. **Simple random sampling**: This is where every subject or respondent has the same chance of being selected. Simple random sampling can be done either with replacement – where subjects are replaced after selection from the entire population or without replacement – whereby subjects are withdrawn from the entire populations after being selected.

   For a reliable and practical sampling, it is advisable to apply a simple random sampling without replacement as this is capable of improving the chances of subjects yet to be part of the sample or yet to be selected.

2. **Systematic sampling**: This involves systematic selection of subjects from a given population. In this selection process with a starting point (or subject) selected randomly without replacement, the required sample size is selected using a sampling interval**.**

This sampling interval is obtained by the formula:

Sampling Interval = $\frac{N}{n}$,

Where N = size of population
n = the sample size.

For example,

Let the serially listed subjects in the population be N = 20, and the required sample size be, n = 5. Then, the sampling interval will be

$N/n = 20/5 = 4$.

The listings are as follows:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10
11, 12, 13, 14, 15, 16, 17, 18, 19, 20.

Assuming that we randomly select the 7^{th} listing, then starting from the 7^{th} listing, every 4^{th} subject will be selected as part of the sample as follows:

7, 11, 15, 19, and 23.

Thus, subjects with serial numbers 7, 11, 15, 19, and 23 will be selected

3. **Stratified sampling**: This gives a fair representation of various strata in a given population. To illustrate this technique, assume a working population of 500 executives with:

10 General Managers
100 Managers
200 Assistant Managers
190 Supervisors

To obtain a representative stratified sample from this population, the stratification is a follows:

	Number	Percent
General Managers	10	2%
Managers	100	20%
Asst. Managers	200	40%
Supervisors	190	38%
Total	500	100%

Assume we want to obtain a stratified sample of 100 executives from the population, the sample will be made up of:

General Manager	= 2% of 100 =	2
Managers	= 20% of 100 =	20
Asstistant Managers	= 40% of 100 =	40
Supervisors	= 38% of 100 =	38
	Total	100

From this process, we would have a representative sample of 2 General Managers, 20 Managers, 40 Assistant Managers and 38 Supervisors.

4. **Cluster sampling**: This sampling technique is useful especially where ethnic or geographical representation is needed in a given study. In this technique, the researcher selects a geographical area at random. Every single subject in such area is then used as part of the sample.

   For all practical purposes, and to ensure minimum bias with statistical analysis and results, it is advisable to use any of the above probabilistic sampling techniques.

3.2.2 Non-Probabilistic Sampling Techniques

These techniques include:

1. **Convenience sampling (or accidental sampling)**: Is applied by researchers interested in having ideas of situations or phenomenon of interest. A marketing researcher for Nigerian Breweries, for example, may decide to station himself at a beer parlor in order to obtain the opinions of consumers of a particular brand of the brewery's beer. The intention is for the researcher to interview any of the consumers he may come in contact with accidentally. It is a good example of convenience or accidental sampling.

   As with any other non-probabilistic sampling design, convenience or accidental sampling design may give rise to unreliable inferences and conclusions. This is however, economical and simple to use.

2. **Judgment sampling:** This is applicable to situations in which the researcher is guided by the belief that reference subjects will provide the required data or information for a particular research process. A researcher interested in studying the economic implications of bank consolidation in Nigeria, for example, is not likely to get reliable information from petty traders. Using his or her own judgment, the researcher would choose respondents from the banking industry or profession as this class of respondents would be in a better position to provide the relevant information on bank consolidation.

3. **Quota sampling:** This is similar to stratified sampling technique as it selects representative subjects according to their percentage representation in the population. It is mostly used in cases

whereby the characteristics of the population of interest can be easily identified. These characteristics are usually represented in the sampling process so the basic information about the population can be obtained. A typical example of quota sampling would be in the case of a population consisting of different groups such as, students, lecturers, and parents. In such a population, the researcher may want to classify the population according to size as follows:

Students	………………	60%
Lecturers	………………	20%
Parents	……………...	20%
Total	……………..	100%

Assume the researcher is interested in a sample size of 5000 subjects, then, using the quota sampling design, he or she would choose the sample of students, lecturers, and parents as follows:

Students	--------------	60% of 5000	=	3000
Lecturers	……………	20% of 5000	=	1000
Parents	……………	20% of 5000	=	1000
Total	……………	100%		5000

Non-probabilistic sampling techniques are not advisable in practical research.

3.3 Determination of the Sample Size

Techniques for the determination of the sample size for a given research depend on the type of research as well as the size of the population under investigation. For most practical business studies, the population under investigation is usually of a finite or known size, N. In this case, the following formula, as advanced by Taro Yamane, has been suggested for use for determination of sample size.

$$n = \frac{N}{1+Ne^2}$$

where n = desired sample size
N = the finite size of the population
e = maximum acceptable margin of error as determined by the researcher
1 = a theoretical or statistical constant

Example: Consider a researcher working on a population of 310 subjects. He would like to determine the sample size with a five percent margin of statistical error. According to the above formula, the required sample size would be:

$$n = \frac{N}{1 + Ne^2} = \frac{310}{1 + 310(0.05)^2}$$

$$= 175$$

It follows that the required sample size would be approximately 175 subjects

Other formulas that can be used, according to the research needs are discussed below.

3.3.1 Sample Size for Mean Values

The sample size for statistical estimation and tests relating to population mean or average values can be obtained by:

$$n = \frac{Z^2\sigma^2}{e^2}$$

where Z = the Z-statistic (or value) corresponding to the desired confidence level

σ = a pre-determined value of the population standard Deviation

e = the maximum acceptable margin of error

The value for the population standard deviation (σ) can be obtained from previous statistical studies of similar population. If such information is not available, a satisfactory value can be obtained through pilot surveys, using a reasonable sample of the population.

Example: Suppose you want to be 99% confident that the true value of the population mean for the PDP voters in Enugu State will be within 10% of the sample mean. By implication, the estimate of the true population mean by the sample would be in error only 10%. Your look at the normal distribution would reveal that the value of Z corresponding to 99% level of confidence or alternatively, one percent level of significance, is 2.58. If you obtained the information that a survey of similar population indicates that 0.8 is a realistic estimate of the population standard deviation, then, the sample size will be:

$$n = \frac{Z^2\sigma^2}{e^2} = \frac{(2.58)^2(0.8)^2}{(0.1)^2} = 426$$

3.3.2 Sample Size for Proportions

The sample size for statistical estimation and tests of proportions is obtained by:

$$n = \frac{Z^2 pq}{e^2}$$

where p = the approximate value of the proportion of success
q = 1 – p = the estimated true proportion of failure.

Note that p is the population proportion to be estimated by taking the sample. If the value of p can be estimated based on past data or experience, then this value can be applied to the formula. In the absence of past data, you are advised to make use of the value of p that will make the quantity pq as large as possible, and such value is p = 0.50.

Example: Suppose that you as a researcher is interested in knowing the sample size needed to estimate the proportion of your customers who are females. You wish to be 99% confident, with a fairly small sampling error of 2.5% in estimating the true proportion. There are no prior information about the true proportion, p, so that, for a conservative estimate, p = 0.50.

With 99% level of confidence or 1% level of significance, the normal distribution table on Z would indicate the value of Z as 2.58. Using the applicable formula, we get:

$$n = \frac{Z^2 pq}{e^2} = \frac{(2.58)^2 (0.5)(0.5)}{(0.025)^2} = 2663$$

It follows that the required sample size would be approximately 2663 subjects.

SELF-ASSESSMENT EXERCISE

With practical examples, can you demonstrate the differences between a sample and population in your chosen research topic?

4.0 CONCLUSION

The unit exposed you to the differences between a population and a sample. Emphasis was placed on the different sampling techniques, categorised into probabilistic and non-probabilistic techniques. The techniques in the determination of the sample size for a given survey research were also presented.

5.0 SUMMARY

The term population depicts the totality of all behaviours in a given universe. It is synonymous to the universal set of all behaviours. A sample, on the other hand is a subset of a population. Statisticians use sample behaviours to draw inferences about the population behaviour.

There exist two different categories of sampling techniques: (i) Probabilistic sampling techniques, including simple random sampling; systematic sampling; stratified sampling; and, cluster sampling techniques. (ii) Non-probabilistic sampling techniques, including convenience sampling; judgment sampling; and quota sampling techniques.

Three different levels of sample size determination were emphasised including, sample size for finite population; sample size for mean values; and, sample size for proportions. It is the type of statistical information that is required that will determine which level of determination will be employed.

6.0 TUTOR-MARKED ASSIGNMENT

A researcher chose to research on a company with a population of 300 employees made up of:

20 General Managers
45 Assistant General Managers
75 Managers
60 Supervisors
100 Junior Staff

(a) Determine the statistical sample for the research.
(b) Produce a stratified sample of the subjects for the research.

7.0 REFRENCES/FURTHER READING

Chinelo Obeleagu-Nzelibe (ed.). (1995). *Business Statistics: Theory and Applications.* Enugu: Optimal Publishers.

Nachimias, D. & Nachimias, C. (1976). *Research Methods in the Social Sciences*. New York: Martin's Press.

Onwe, O. J. (2007). *Statistical Methods for Business and Economic Decisions: A Practical Approach.* Lagos: Samalice Press.

Onwe, O. J. (2012). *Fundamentals of Managerial Statistics and Econometric Analysis: Some Practical Approaches.* Lagos: Impressed Publishers.