NATIONAL OPEN UNIVERSITY OF NIGERIA

SCHOOL OF EDUCATION

COURSE CODE: EDA 844

COURSE TITLE: STATISTICS FOR EDUCATIONAL MANAGEMENT

Course Guide

Course Code:	EDA 844
Course Title:	STATISTICS FOR EDUCATIONAL MANAGEMENT
Course Developer:	Dr. K. A. Salami Emmanuel Alayande College of Education Oyo
Course Writer:	Dr. K. A. Salami Emmanuel Alayande College of Education Oyo
Course Editor:	Prof. J. K. Adeyemi Faculty of Education, University of Benin, Benin City.
STAFF IN CHARGE	Dr S O Ogundiran: National Open University of Nigeria Victoria Island Lagos
Course Coordinator	Dr S O Ogundiran: National Open University of Nigeria Victoria Island Lagos

CONTENTS

1.0 Introduction

The number of people studying statistics increased after its importance was realized in and immediately after the Second World War. First, it was studied only at graduate level in universities, but later it spread to the undergraduate programmes and finally to all levels of education. Again, the complexity witnessed in industries and education enterprise in the 20th century has now forced education managers and administrators to daily engaged in making far reaching decisions for their organizations. Unfortunately, such decisions are usually marked by uncertainty because of the available information, which in most cases, is either incomplete or inadequate. In such situation, education managers usually resort to some kind of estimation which may be guesswork or something based on experience or rather, something evolved by the use of scientific method. Because of this, managers and captains of industries including the educational system have come to realize that knowledge of statistics becomes necessary and essential for proper decision making.

As a graduate student, this course will expose you to various and necessary aspects of statistics. The course also includes some case studies with the sole aim of teaching the students how statistical analysis could be used for solving managerial problems.

This course is a 2 credit unit course divided into three modules viz:

i) Basic concepts in educational statistics.

ii) Testing of hypothesis and generalization

(iii) Data processing in educational management

Each of these modules contains some study units. For example, module one contains 6 units, module two contains 4 units, while module three has 4 units. You will be expected to go through each of the units carefully attending to all built in self-tests. You will also be expected to submit a tutor-marked assignment after each unit. Most of these assignments involve your use of scientific calculator. This will help you with your future assignments and remember, your assignments are as important as your examinations as they carry equal weightings.

The course guide tells you briefly what the course is all about; what course materials you will be using and how you can work your way through these materials.

The course guide also suggests some general guidelines for the amount of time you are likely to spend on each unit of the course in order to complete it successfully. It also gives you some guidance on your tutor marked assignments. Detailed information on tutor-marked assignments is found in the separate Assignment File, which will be available to you.

2.0 Course Aims and Objective

This course aims to introduce you to educational statistics, its basic concepts, relevance, branches. Its basic tools and their uses in education, probability theory and distribution; others are hypothesis testing and how statistical analysis could be employed to solve management problems in education.

This course aims to:

- ü introduce you to the basic concepts in statistics
- ü explain the roles of statistics in education
- ü introduce you to the language of statistics
- ü explain the two measures of central tendency & variability
- ü demonstrate the use of these statistical tools through hypothesis testing
- ü appreciate the contributions of probability theory and distribution to human understanding of chance
- ü justify why the knowledge of statistics has become imperative and a must for education managers and administrators.

To achieve the aims set above, the course sets overall objective. In addition, each unit has specific objectives included at the beginning of a unit. You may want to refer to them during your study of the unit to check on your progress. You should always look at each unit objectives after completing a unit to be sure you have done what was required of you in the unit. Set out below is the wider objectives of the course as a whole. By meeting these objectives you should have achieved the aims of the course as a whole.

On successful completion of the course you should be able to:

- ü advance reasons for the role and relevance of statistics in educational management.
- ü understand the language of statistics as different from mathematics.

- ü differentiate between measures of central tendency and measures of variability.
- ü appreciate that raw numbers by themselves are not useful until they are put into statistical analysis.
- ü list the four major tools commonly used in education statistics, i.e. Correlation coefficient, Chi-Square statistic, Student t -statistic and Regression analysis
- ü demonstrate how to use all these statistical tools for management decision making in education.
- ü state the need for hypothesis testing with a small sample with the aim of determining population related problems.
- ü trace the history of some statistical tools used in this course.
- ü relate the probability theory and distribution to human chances of living or death, successes or failure and why every event in human life is based on probability theory.
- ü justify why every education managers and administrators must learn how to use statistical analysis for decision making.
- ü appreciate that chance and luck play an important role in human lives.

3.0 Main Content

3.1 Course Materials

Major components of the course are:

- ü Course Guide
- ü Study units
- ü Journals and textbooks
- ü Assignment files
- ü Presentation schedule

3.2 What You Will Learn in This Course

As a postgraduate student in this course, you must have been exposed to basic or elementary statistics at the undergraduate level or at the Polytechnic or at the Colleges of Education level. Therefore the overall aim of EDA 844: Statistics for Educational Management is to introduce you to a more advanced but simplified level of statistics. Particularly, a functional statistical analysis needed for decision making

in organizations. Essentially, you will find the unit dealing with hypothesis testing more interesting and rewarding as it will teach you how statistical analysis could help you make management decisions.

3.3 Study Units

There are fourteen units in this course as follows:

MODULE ONE

Unit 1: Basic Concepts in Educational Statistics:

Unit 2: The Role of Statistics in Educational Management

Unit 3: Basic Statistical Vocabularies, Notations and Symbols

Unit 4: Frequency Distribution

Unit 5: Measures of Central Tendency

Unit 6: Measures of Variability

MODULE TWO

Unit 1: The Probability and Non- Probability Sampling

Unit 2 The Probability Theory and Distribution

Unit 3: Estimation

Unit 4: Testing of Hypothesis

MODULE THREE

Data Processing in Educational Management

Unit 1: The Correlation

Unit 2: The CHI SQUARE (X^2)

Unit 3: The Student t Statistic

Unit 4: **The** Regression Analysis

The first six units constitute module one, which is on the basic concepts in statistics. The next four units constitute module two and this is on testing of hypothesis by using the statistical tools in module three. The next four units constitutes module three, which is on data processing in educational management. Here in this module, relevant statistical tools such as correlation, Chi-Square, Student t statistic and both simple and multiple regression equations and analysis were treated. Those tools are necessary working tools with which education managers could use and make

decisions concerning education matters. Each of these units is designed in such a manner that it would not take you for more than a maximum of two and half hours. However, when you are expected to consult some texts and perhaps make use of your scientific calculators for your tutor marked assignment, you may likely spend some extra 30 minutes. But as a research student you have to learn to do this within a stipulated time.

The final examination covers information from all parts of the course.

3.4 Course Marking Scheme

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

Assessments	Marks
Assignment 1 15	15 assignments, the best 6 out of
Final Examination	15 will be picked. 6 x 5 = 30%
Total	70% of overall course marks
	100% of course marks

3.5 Course Overview

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

Unit	Title of Work	Weeks	
	Assessment	*Activity*	*End of Unit*
1.	Basic Statistics in Educational Management	3	1
2.	The Role of statistics in Educational Management	3	2

3	Basic statistical vocabularies, Notations and Symbols	2	3
4	Frequency Distribution	4	4
5	Measures of Central Tendency	4	5
6	Measures of Variability	4	6
7	Data Processing in Educational Management 1: (The Correlation)	2	7
8	Data Processing in Educational Management 2: (The CHI SQUARE)	2	8
9	Data Processing in Educational Management 3: (The Student t statistic)	2	9
10	Data Processing in Educational Management 4: (The Regression Analysis)	2	10
11	Probability and Non- Probability Sampling	2	11
12	Probability Theory and Distribution	3	12
13	Estimation	2	13
14	Testing of Hypothesis	4	14

3.6 How to get the best from this Course

In distance learning the study units replace the university lecturer. This is one of the great advantages of distance learning. You can read and work through specially designed and prepared materials at your own pace and at a time and place that suit you best. Think of it as you read the lecture instead of listening to a lecturer.

In the same way that a lecturer might set you some reading to do, the study units tell you when to read your set books or other materials, and when to undertake computation work. Just as a lecturer might give you an in-class exercise, your study units provide copious exercises for you to do at appropriate points.

Each of the study units follows a common format. The first item is an introduction to the subject matter of the unit and how a particular unit is integrated with the other units and the course as a whole. Next is a set of learning objectives.

These objectives allow you to know what you should be able to do by the time you have completed the unit. You should use these objectives to your study. When you have finished the unit, you must go back and check whether you have reached the objectives. If you make a habit of doing this, you will significantly improve your chances of passing the course.

The main body of the unit guides you through the required reading from other sources. This will usually be either from your set books or from a reading section. Some units require you to undertake some working with your scientific calculators or necessitated you to understand some symbols and notations. You will also be directed when you need to use a computer. The purpose of computation, using scientific calculators and learning by heart is two-fold. First, it will enhance your understanding of the basic principles of the material unit. Second, it will give you practical experience of using programmes, which you could well encounter in your work outside your studies. In any event, most of the techniques you will study are applicable on computers in normal working practice, so, it is important that you encounter them during your studies.

Activities in form of self-tests are interspersed throughout the units, working through them will help you to achieve the objectives of the unit and prepare you for the assignments and examinations. You should please do each self-test as you come across it in the study unit. There are many examples given in the study units; work through these several times when you come across them too. In fact, these examples are deliberately provided so that working through them several times will give you a firm grip over them and self-test that may follow.

The following is a practical strategy for working through the course. If you run into any trouble, telephone your facilitator or post the question (s) to him. Remember that your facilitator s job is to help you. When you need help which is obvious with this course, don t hesitate to call and ask your facilitator to provide it.

3.7 How to organize your study

3.7.1 **Organize a study schedule**. Refer to the course overview for more details. Note the time you are expected to spend on each unit and how the assignments relate to the units. More importantly, details of your tutorials, and

the date of the first day of the semester will be made available to you. You need to gather together all this information in one place, preferably in your diary or a wall calendar or in your handset. Whatever method you choose to use, you should decide on and write in your dates for working on each unit.

3.7.2 Once you have created your own study schedule, do everything you can to stick to it. The major reason that makes students fail is that they get behind with their course work. If you get into difficulties with your schedule, please let your facilitator knows before it is too late for help.

Ø Turn to unit one and read the introduction and the objectives of the unit

Ø Assemble the study materials. Information about what you need for a unit is given in the overview at the beginning of each unit. You will always need both the study unit you are working on and one of your set books on your desk at the same time.

Ø Work through the unit. The content of the unit itself has been arranged to provide a sequence for you to follow. As you work through the unit, you will be instructed to read sections from your set books or other articles. Use the unit to guide your reading.

Ø Keep an eye on the course information that will be continuously posted to you by the university.

Ø It is advisable before the relevant due dates (about 4 weeks before due dates) to take the Assignment File and your next required assignment. Keep in mind that you will learn a lot by doing the assignments carefully. They have been designed to help you meet the objectives of the course and, therefore, will help you pass the examination. Submit all assignments not later than the due date.

Ø When you have submitted an assignment to your facilitator for marking, do not wait for its return before starting on the next unit. Keep to your schedule. When an assignment is returned, pay particular attention to your tutor s comments, both on the facilitator marked assignment form and

also the one written on the assignment. Consult your facilitator as soon as possible if you have any question or problem.

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

3.8 Tutorial Sessions

There are 20 hours of tutorials (ten 2 hours sessions) provided in support of this course. You will be notified of the dates, time, and location of these tutorials, together with the name and phone number of your facilitator, as soon as you are allocated a tutorial group.

Your facilitator will mark and comment on your assignments, keep a close watch on your progress and on any difficulties you might encounter and provide assistance to you during the course. You must mail your tutor-marked assignments to your facilitator well before due dates (at least two working days are required). They will be marked by your facilitator and returned to you as soon as possible.

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

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

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

3.9 Assignment Files

There are twenty assignments in the course. Units one, two, eleven and sixteen have two assignments each while others have one assignment each. As for those assignments that were included, you will find all the details of the works you must submit to your facilitator for marking. Remember that those assignments are as important as the examinations as they carry equal weightings.

3.10 Assessment Procedure

There are two aspects to the assessment of the course. First, are the facilitator-marked assignments; second, is the written examination. In tackling the assignment, you are expected to apply information, knowledge and techniques gathered during this course. The assignments must be submitted to your tutorial facilitator for formal assessment in accordance with the deadlines stated in the presentation schedule and the assignment files. The work you submit to your facilitator for assessment will count for 30% of your total course work.

At the end of the course, you will need to sit for a final written examination of not more than three hours duration. This examination will also count for 70% of your total course mark.

3,11 Final Examination and Grading

The final examination for EDA 844 will not be more than three hours duration and have a value of 70% of the total course grade. The examination will consist of questions which reflect the types of self-testing, practice exercises and tutor marked problems you have previously encountered. All areas of the course will be assessed. Use the time between finishing the last unit and sitting for the examination to revise the course. You might find it useful to review your self-tests, tutor marked assignments and comments on them before the examination.

4.0 Conclusion.

This course is a 2 credit unit course divided into three modules viz:

(i) Basic concepts in educational statistics.

(ii) Testing of hypothesis and generalization

(iii) Data processing in educational management

Each of these modules contains some study units. For example, module one contains 6 units, module two contains 4 units, while module three has 4 units. You will be expected to go through each of the units carefully attending to all built in self-tests. You will also b e expected to submit a tutor-marked assignment after each unit. Most of these assignments involve your use of scientific calculator. This will help you with your future assignments and remember, your assignments are as important as your examinations as they carry equal weightings.

The course guide tells you briefly what the course is all about; what course materials you will be using and how you can work your way through these materials. The course guide also suggests some general guidelines for the amount of time you are likely to spend on each unit of the course in order to complete it successfully. It also gives you some guidance on your tutor marked assignments. Detailed information on tutor-marked assignments is found in the separate Assignment File, which will be available to you. Again, it exposes you to the general format you will come across throughout this course

5.0 Summary

EDA 844 intends to introduce you to educational statistics, the basic knowledge and principles of statistics necessary for decision making in the management of educational system.

Among others, you will be able to answer these kinds of question.

- What is the definition of statistics?
- What is the relevant of statistics in education?
- Is it true that statistics have vocabularies, notation and symbols?
- Is it possible to get general answer to population problem from a small sample?
- Is it true that numbers by themselves are useless until they are put into statistical analysis?
- What are the major statistical tools needed for management decision making?

- Is it true that life and human living is based on probability?
- How can we generate acceptable decision from hypothesis testing?
- To what extent are probability theory and distribution of immense help to education managers and administrators?
- What are the merits and limitations of statistics in decision making?
- Is it possible to use or employ statistical analysis to answer or solve managerial problem? If yes, how?

6.0 Teacher Marked Assignments (ATMs)

Assignment questions for the units in this course are contained in the Assignment File. You will be able to complete your assignments from the information and materials contained in your set books, reading, studying units and probably the internet. However, it is desirable in all degree level education to demonstrate that you have read and researched more widely particularly in statistics, than the required minimum.

Using other texts will give you a broader viewpoint and will definitely provide a deeper understanding of the subject.

When you have completed each assignment, send it, together with your TMA (tutor marked assignment) form, to your facilitator. Make sure that each assignment reached your facilitator on or before the deadline given in the presentation schedule and Assignment File. If, for any reason, you cannot complete your work on time, contact your facilitator before the assignment is due to discuss the possibility of an extension. Extension will not be granted after the due date unless there are exceptional circumstances. There are thirty eight (38) tutor marked assignments in this course. You only need to submit 15; out of which the best 6 will be picked. Each of these 6 carries 5 marks each making up your 30% continuous assessment.

7.0 References/ Further Readings

Adewoye, S. O. (2004): Basic Statistics for Engineering, Economics and Management, Lagos: Olukayode Ojo Commercial Enterprises.

Clark, G. M. and Cooke, D. (2004). Basic Course in Statistics 5th ed. New York. Oxford University Press Inc.

Hildebrand, D.K. & Lyman, O. (1998). Statistical thinking for managers (4th Ed.) California: Duxbury Press.

Levin, R.I. &Rubin, D.S. (1997). Statistics for Management (7th Ed.) New Jersey: Prentice Hall International.

Monga, G.S. (2007). Mathematics and Statistics for Economics (Second Revised Edition) New Delhi. VIKAS Publishing House PVT Ltd.

Salami, K.A. (1999). Descriptive Statistics for Beginners. Oyo Odumatt Press and Publishers.

Salami, K. A. (2002). Statistical Models and Projections in Educational Management. Lagos: Master Printers International.

Journals

Salami, K.A. (2001). Introduction to Basic Statistics in Management. In Adeyanju, A. (Ed.) Introduction to Educational Management. Oyo Green Light Press and Publishers.

Salami, K. A. (2001). Basic statistics and Data Processing in Educational Management in A. Adeyanju (ed) Introduction to Educational Management, Oyo, Green Light Press and Publishers.

Salami, K. A., Oyeniran, J. O., Adebiyi, M. E. (2004). Perceptions of Employers of Labour on the quality and proficiency of sandwich degree graduates in Oyo State, Nigeria Ado-Ekiti Journal of Educational Foundation and Management. Vol. 1 No. 1&2 pp 97 108.

Salami, K. A., Raji, R. A. (2006). Perceptions of Lagos State Employers of Labour on the quality of sandwich degree graduates in the labour market . The Pacesetter Vol. 13 No. 1 pp 315 326.

Course Code:	EDA 844
Course Title:	STATISTICS FOR EDUCATIONAL MANAGEMENT
Course Developer:	Dr. K. A. Salami Emmanuel Alayande College of Education Oyo
Course Writer:	Dr. K. A. Salami Emmanuel Alayande College of Education Oyo
Course Editor:	Prof. J. K. Adeyemi Faculty of Education, University of Benin, Benin City.
STAFF IN CHARGE	Dr S O Ogundiran: National Open University of Nigeria Victoria Island Lagos
Course Coordinator	Dr S O Ogundiran: National Open University of Nigeria Victoria Island Lagos

MODULE ONE

Unit 1 Basic Concept in Educational Statistics: Definition, Characteristics, Relevance, Scope and Branches.

CONTENTS

1.0 Introduction

This unit will introduce you to various definitions of statistics as perceived by different experts. The roles of statistics as well as its relevance in education are adequately highlighted. Like other academic subjects, statistics also has its own characteristics, branches and language. Both of these equally discussed with examples where necessary

.

2.0 Objectives

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

(i) define certain concepts in statistics.

(ii) outline major characteristics and functions of statistics

(iii) identify the various branches of statistics.

(iv) explain the scope of statistics

3.0 Main Content

3.1 Definition of Statistics

A Marxist defined statistics as an independent social science which studies the quantitative aspect of mass social phenomena, regularities of social development, social production and the influence of natural and technical factors on quantitative changes in social life.

Secrist on the other hand defined statistics as aggregates of facts affected to a marked extent by a multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other.

Salami (1999) however, defined statistics as the scientific method of collecting, organizing, analyzing, summarizing, presenting, collating and interpreting data.

Monga (1972) identified statistics as a science as it describes facts objectively without delivering value judgement. It is an art in that it applies certain techniques to available facts, sometimes in a subjective manner. Being a systematic body of knowledge, statistics is considered a science, though not one as exact as physical sciences.

Kerlinger and Howard (2004) defined statistics as the theory and method of analyzing quantitative data obtained from samples of observations in order to study and compare sources of variance of phenomena, to help make decisions; to accept or reject hypothesized relations between the phenomena; and to aid in drawing reliable inferences from empirical observations.

Adewoye (2004) opined that statistics is concerned with scientific methods of collecting, organizing, summarizing, presenting and analyzing data; as well as drawing conclusions and making reasonable decisions on the basis of such analysis.

Based on these definitions, the word statistics may be used either in the plural or singular form. For example:

(i) Statistics are numerical facts systematically collected (plural).

(ii) Statistics is the science dealing with the collection, organisation, analysis and interpretation of numerical data. It is a systematic body of knowledge (singular).

Statistics as Singular and as Plural

It is very important for education managers to note that the word statistics could be used as singular or plural. According to Adewuyi and Oluokun (2001), statistics when used as singular, always referred to a specialized human activity which is concerned with the collection, ordering, analyzing and interpretation of data. The authors believed that using statistics as singular primarily refers to the scientific method of collecting, organizing, presenting and analyzing data.

However, the authors submitted that the plural form of statistics refers to the systematic collection of numerical data about some events or subjects. It may denote the data themselves or numbers. The most important thing is for the education manager to be consistent and familiar with the form in which he uses the concept.

3.2 The Place of Statistics in Education

Educational system in the 21st century has grown into a complex industry full of challenges and abstract ideas. Therefore, the immediate reaction of modern man or education manager of today to the apparent complexity of the world around him has been to formulate for himself a simplified model of that complex world. He then tries to substitute this cosmos of his own for the world experience. Whereas, in the idealized world including the system of education, the real system is always decentralized into a series of simplified system so that the characteristics required for the solution of a problem become conspicuous. It is this complexity in the present educational system and how to simplify them that brings about idea of statistics, hence the prominent place given to statistics in every day human activities.

3.3 Relevance of Statistics to Educational Management

Education managers frequently rely on inputs from statistical analysis to help them make decisions and undertake some planning exercise. This is because, management problems usually emanate from the real world problems. In the case of educational system, real world problems may include insufficient supply of relevant textbooks, shortage of physical resources like chairs, benches, desk, chalk, classrooms, office accommodation, or personnel etc. To solve this kind of problem nation wide, the education manager needs the services of statistics in projecting the resources needed.

Another relevance of statistics in education is when we translate our desire or wants into statistical equation .

Example 1.1. Let us examine the model of a quadratic equation stated in words such as if the product of five and the square of unknown quantity is added to the product of seventeen gives the result as twelve .

Example 1.2. In the case of equilibrium principles in economics, If the quantity demanded of commodity X equals two hundred minus two multiplied by unknown variable and that of supply equation is given as one hundred plus unknown variable.

To find solution to these two examples become a little difficult even to the brilliant brains, because it may be difficult to determine the unknown variables mentioned in the examples. Perhaps, what the best brain will resort to is what could be called Trial and Error method . But such a method may not serve when the model or equations are complex, and the quantities are in decimal. Again when we rely on sentence models only, creative and quantitative planning would be much limited thereby hindering development.

Statistical Solution

To see the beauty of statistics in education, let us recast the same models in statistical form.

The first equation will be

$$5x + 17x^2 = 12$$

While the second equation will be

$$Qd = 200 \quad 2p$$

$Qs = 100 + p.$

It is now very easy with these equations to determine the unknown variables.

3.4 Characteristics and Functions of Statistics

For ease of identification, characteristics here are illustrated with italics while the functions are expressed in prose.

i. ***Numerical expression:*** Statistics deals with quantitative data, with numbers as expressions of meaningful relationship. Statistical data are not abstract numbers. They constitute concrete material which represents objects and their measurements.

ii. ***Accuracy:*** The degree of accuracy depends on circumstances. Based on the nature and object of an inquiry, a reasonable standard of accuracy has to be maintained.

iii. ***Systematic collection of data:*** There should be a definite method and purpose in the collection of data. A systematic procedure should be followed.

iv. ***Aggregates of facts:*** Isolated figures are not statistics. Masses of facts in social sciences contain the laws of general behaviour. Statistical methods extract meaningful information from a mass of data.

v. ***Multiplicity of causes:*** There are various unassignable causes acting and reacting with one another when we consider any phenomenon.

vi. ***Collection of data for a predetermined purpose:*** A statistician should proceed with a specific object in view. Vagueness and ambiguity of purpose can lead to waste of time, energy and money.

vii. ***Comparison of data:*** Relationships between and comparison of variables are an important part of the study of statistics.

viii. ***Hypothesis testing, prediction and formulation of policies:*** These are parts of any scientific study. Statistical methods are useful in formulating and testing hypotheses, forecasting future events and framing suitable policies.

Self Assessment Exercise 1

Mention and discuss four characteristics and functions of statistics.

3.5 Branches of Statistics

As a result of historical development, statistics may be classified into two parts, namely: descriptive statistics and inferential or analytical statistics.

3.5.1 Descriptive Statistics

Descriptive statistics involves the collection, analyzing and informative presentation of a mass of numerical data. Such data may be quantitative such as measures of height and weight, or qualitative such as sex and personality. Descriptive statistics would not present information more than what we can touch, see o r measure. Its function is basically to describe the event or outcome without drawing any conclusion. Statisticians regard descriptive statistics as data analysis without probability. Attributes such as the mean, mode, median, geometric mean, harmonic mean, range, mean deviation, standard deviation, percentile, kurtosis, proportion and correlation co-efficient etc fall within the domain of descriptive statistics.

3.5.2 Inferential Statistics

Inferential statistics is a formalized body of techniques for making conclusions concerning the properties of a large collection of data from an examination of a sample of the collection. Its purpose is to surmise the properties of a population from knowledge of the properties of only a sample of the population.

The inferential statistics builds upon descriptive statistics by making interpretation there-from with the powerful tool of probability theory. Inferential statistics enables education managers to make decisions, generalization, predictions and conclusions with accuracy and the validity of such conclusions can be ascertained any time. The scope of inferential statistics includes, operation research, linear programming, games theory, t-test, f-test, simple and multiple regression and other test of significance.

Self Assessment Exercise 2

With examples discuss the relevance of statistics in education.

4.0 Conclusion

In this unit, students have been acquainted with definitions of statistics, major functions of statistics as well as identification of two main branches of statistics, which are descriptive and inferential statistics. In addition, students have been exposed to various fields where statistical tool are useful.

In the next unit, students will be introduced to the major roles of statistics in educational management.

5.0 Summary

In this unit, students have gained insight into the

i. definition of statistics
ii. identification of the two main branches of statistics
iii. understanding the job of statistics as different from arithmetic and mathematics
iv. Identification of the various fields where statistics could be used successfully.

6.0 Tutor Marked Assignment

(a) Highlight four characteristics and functions of statistics.

(b) Discuss the major differences between the descriptive and inferential statistics.

7.0 References/ Further Readings

Adewoye, S. O. (2004): Basic Statistics for Engineering, Economics and Management, Lagos: Olukayode Ojo Commercial Enterprises.

Hildebrand, D.K. & Lyman, O. (1998). Statistical thinking for managers (4th Ed.) California: Duxbury Press.

Levin, R.I. &Rubin, D.S. (1997). Statistics for Management (7th Ed.) New Jersey: Prentice Hall International.

Monga, G.S. (2007). Mathematics and Statistics for Economics (Second Revised Edition) New Delhi. VIKAS Publishing House PVT Ltd.

Salami, K.A. (1999). Descriptive Statistics for Beginners. Oyo Odumatt Press and Publishers.

Salami, K.A. (2001). Introduction to Basic Statistics in Management. In Adeyanju, A. (Ed.) Introduction to Educational Management. Oyo Green Light Press and Publishers.

UNIT 2 THE ROLE OF STATISTICS IN EDUCATIONAL MANAGEMENT

CONTENT

2.0 Introduction

This unit will teach you that in educational management, planning, organizing, supervision, control and decision making are necessary at every stage. And that at all the stages, an analysis of data collected becomes easier with the help of statistical methods. This unit will illustrate to you the relationship between management problems and statistical solution through a flow diagram as well as a case study.

2.0 Objectives

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

i. identify the relationship between management problems and statistical solution to such problems.

ii. trace the process of interaction between the management problems and statistical solution from a flow diagram.

iii. appreciate how statistical analysis can solve managerial problems through a case study presented.

iv. identify how statistics could be misused.

v. discover the limitations of statistics.

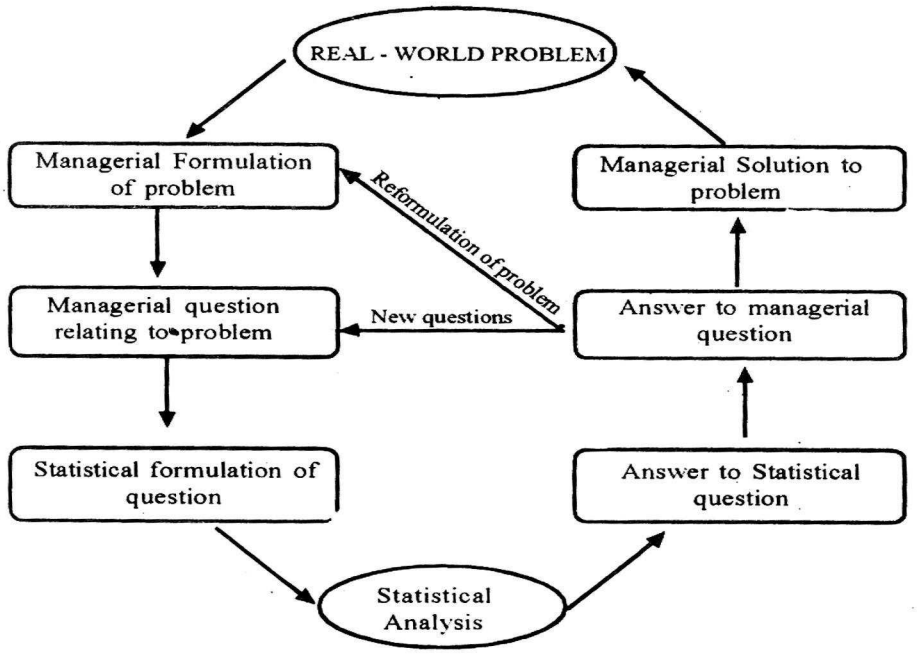
3.0 Main Content

3.1 The Role of Statistics in Educational Management

As mentioned earlier in the introduction of this unit, the management of education involves planning, organizing, supervision, control and decision making at every stage. And that at all these stages, an analysis of data collected becomes easier with the help of statistical methods. In the same vein, managers of education frequently rely on inputs from statistical analysis to help them make decisions and perhaps undertake some reasonable planning. This is because management problems usually emanate from the real world situation. For example, education problem in the real world could be insufficient supply of relevant textbooks, short supply of physical resources, fall in the standard of knowledge and skills imparted by the teachers and acquired by the learners etc.

To solve this kind of management problems, an education manager would need the services of statistics in calculating and establishing the facts of the case. To do this, the education manager will collect sample data for analysis which is expected to answer the statistical questions raised. Therefore, the manager needs the knowledge of statistics.

This process is illustrated with a flow diagram in figure 1 below.



Flow diagram showing the role of statistics in managerial decision making

The flow diagram started with the management of education in the real world. As a result of feed-back from the society about the education products, the manager can now formulate a problem to be investigated. From here the management will put the problems into questions to be answered. The statistician will change the management questions into statistical questions and from sample data collected, carry out statistical analysis. This analysis is to answer the statistical questions formulated which invariably will answer the managerial problems and questions and put the problems to rest or the analysis may suggest a reformulation of the original problem or suggest a new managerial question.

Self Assessment Exercise 1

Give cogent reasons why education managers and administrators need the knowledge of statistics and statistical analysis for making management decisions.

3.2 Purpose of Building Statistical Models in Education

If statistical models are replicas, miniatures and dummies, then, the purpose of statistical models in education would be to simplify and to abstract only those features of reality, which are presumed to be relevant to the educational problems on hand. In other word, the main reason for building statistical model is to reduce the complex world into a manageable unit that could be handled within the four walls of a classroom.

The most fundamental feature of statistical model, according Owolabi (2001) is that their construction involves a highly selective attitude to information by eliminating incidental details, some fundamentals, interesting or relevant aspects of real world appear in some generalized form. Statistical models are therefore constructed to:

(i) simplify the complex world of reality.

(ii) derive pleasure from making them.

(iii) demonstrate the facts, which they illustrate, and

(iv) form part of a permanent collection of similar constructions.

3.3 The Scope of Statistics

Originally, the science of statistics was concerned with figures regarding the resources of a state, man, land and wealth. Nowadays, statistics is used in many fields

of enquiry for describing and analyzing large group of aggregates, which are too complex to be intelligible by simple observations. Statistics is particularly useful for making and/or facilitating comparisons. Today statistics is used in the following spheres.

(i) **Population Statistics**

The census of population gives information about the number of people, their age, occupations, civil conditions, housing conditions, sex, educational background, number in a family etc. From this, education manager can obtain by comparison with the results of previous census enquires, information about the growth of population as a whole, migration situation from rural areas to urban centres, number of additional birth etc. To do all these things successfully require the services of statistics as a tool.

(ii) **Trade and Production Statistics**

It is statistical figure that will provide us with information about a country s exports and imports, the bulk and value of manufactured commodities, the amount of livestock, the produce of fishing industries, agricultural products, the number of ships entering and leaving the ports and such other matter, all of which is of great importance, especially for purposes of taxation and planning.

(iii) **Medical and Vital Statistics**

It is the work of statistics to provide details concerning the number of persons having infectious diseases, how an epidemic has spread and progressed, death rate, birth rate, number of HIV/AIDS victims, accident victims rate every month. By comparing such information with the previous years or months, medical experts would be able to know the level of progress made so far in a country.

(iv) **Social Statistics**

Social statistical figures will show under what condition or circumstances typical men and their families live, what they earn for a living, what rents they pay, what they spend on various items, what type of houses they live in, what type of social amenities they enjoy etc.

(v) **Educational Statistics**

This type of statistical figure will give information about the number of schools in a state or country or locality. The figure will show the number of primary,

secondary, post-secondary institutions like universities, polytechnics, colleges of education, monotechnics etc. It will also indicate number of available classrooms, teachers, learners, laboratories, equipment; student-teacher ratio and so many information about an educational system.

(vi) **Business Statistics**

Statistical figures in this area cover a wide and ever-increasing field. There is hardly any business problem which can not be reasonably answered by collection and analysis of statistics. For example, the daily results of a salesman will show whether or not he is more successful than his colleagues. Production of crude oil last year and this year when compared will give the planners and policy makers a guide for future planning.

3.4 Misuse and Distrust of Statistics in Education

There are many ways in which statistics are likely to be used improperly and wrongly. Some of these ways are listed below:

(i) **Errors of context:** Facts, otherwise true, may be quoted or presented out of context in such a way as to misrepresent the real state of affairs. This is common with politicians, journalists and law enforcing agencies.

(ii) **Errors of Generalization:** A generalization or conclusion based on incomplete or inadequate or unrepresentative data can lead to wrong inferences. For example, on the basis of very poor marks obtained b y two or three students, we cannot conclude that all the students from a university are equally bad.

(iii) **Errors of Deduction:** A general method may be wrongly applied to a specific case. For example, if the students of a particular class have been showing good results in the past years, it does not mean they will necessarily do so in the present or future years..

(iv) **Bias:** Conscious or unconscious entry of bias in statistical work is common. It may enter consciously when it is associated with a purpose. Bias may also come into statistical analysis through exaggeration of quantities or through the use of colourful high sounding words.

(v) **Errors of non-comparability:** Wrong, improper, meaningless or untimely comparisons can lead to faulty conclusions. A comparison of index numbers of

different regions based on different years or different commodities can be meaningless.

(vi) **Errors of over-simplicaiton:** Sweeping statements, omission of essential details and quick half truth are misleading. For example, a mention of a rise in national income without reference to a rise in population cannot give a correct idea of changes in per capita income.

(vii) **Errors of Spurious Accuracy:** As against oversimplication, there may be found an irrelevant attempt at showing high accuracy. To speak of Mr. A s income running into millions of Naira correct to the last kobo does not make much sense.

(viii) **Errors of calculation:** Mistakes in the calculation of ratios, percentages and in the application of mathematical operations are common in statistics and should be avoided.

(ix) **Errors of Assuming Causation:** The assumption of a wrong cause and effect relationship may lead to nonsensical results.

(x) **Abuse:** Abuse of statistics is common. It is often the result of ignorance rather than design. Differences in definitions, methods, sampling procedures, sources of data can give rise to differences in results leading to misunderstanding and misrepresentation.

(xi) **Distrust of Statistics can Result from the Abuse of Statistics and Statistical Concepts:** It is very important to say here that there is nothing wrong with statistical tools. The fault, if any, lies with the user of the science and not with the science.

3.5 Limitations of Statistics

As good as statistics are in assisting education manager to solve educational problems, it has some limitations:

1. that statistics deals with aggregates. Individual facts do not constitute statistics. We need to have a sufficiently large amount of data to study and draw conclusions.
2. statistical results are true only on an average. This is not universally true. To get good results an experiment may have to be repeated several times.

3. statistics can be misused. It is possible to suppress some facts and emphasize others or to mis-represent some parts and represent others partially. This may be done to establish misleading conclusions.

4. some problems may not be amenable to statistical analysis, either because it may be difficult to do so or it may be because of ignorance on the part of the analyst, inadequacy of tools or inappropriate data

4.0 Conclusion

In this unit, students have been introduced to the process of converting managerial problems and questions into statistical problems and questions. The unit has also revealed to the students how statistical analysis can assist the education manager to get solution to the real world problem.

5.0 Summary

In this unit, students have been exposed to:

- Ø the role of statistics in educational management
- Ø the process necessary for the conversion of managerial problems and questions into statistical problems and questions through a flow diagram.
- Ø a case study in which statistics was used to answer the managerial problems.

6.0 Tutor Marked Assignment.

a. Mention and discuss five ways by which statistics could be misused.

b. Enumerate the purpose of statistical model in education

References/ further Readings

Adewoye, S. O. (2004): Basic Statistics for Engineering, Economics and Management, Lagos: Olukayode Ojo Commercial Enterprises.

Hildebrand, D.K. & Lyman, O. (1998). Statistical thinking for managers (4th Ed.) California: Duxbury Press.

Levin, R.I. &Rubin, D.S. (1997). Statistics for Management (7th Ed.) New Jersey: Prentice Hall International.

Monga, G.S. (2007). Mathematics and Statistics for Economics (Second Revised Edition) New Delhi. VIKAS Publishing House PVT Ltd.

Owolabi, S.O. (2001). Statistical Models and Projections in Educational Management. In P.O. Okunola (ed). Theory and Practice of Educational Management. Oyo. OYSCE Publication.

Salami, K.A. (1999). Descriptive Statistics for Beginners. Oyo Odumatt Press and Publishers.

Salami, K.A. (2001). Introduction to Basic Statistics in Management. In Adeyanju, A. (Ed.) Introduction to Educational Management. Oyo Green Light Press and Publishers.

Salami, K. A., Oyeniran, J. O., Adebiyi, M. E. (2004). Perceptions of Employers of Labour on the quality and proficiency of sandwich degree graduates in Oyo State, Nigeria Ado-Ekiti Journal of Educational Foundation and Management. Vol. 1 No. 1&2 pp 97 108.

Salami, K. A., Raji, R. A. (2006). Perceptions of Lagos State Employers of Labour on the quality of sandwich degree graduates in the labour market . The Pacesetter Vol. 13 No. 1 pp 315 326.

Appendix A

3.6 Statistical Solution to Managerial Problem: A Case Study

Introduction

University of Ado-Ekiti, Ekiti State, Nigeria established an outreach centre, known as Sandwich Degree Programme Centre at Emmanuel Alayande College of Education, Oyo in 1997. This distance learning centre like others established elsewhere by the university of Ado-Ekiti, has turned out hundreds of university graduates in various fields of learning including graduates of educational management. Majority of these graduates were on permanent jobs while few of them were unemployed before undertaking the programme.

Real World Problem

After Ten (10) years of the centre s operations (1997-2006), a managerial problem emanated from the society as to the quality of knowledge and skills acquired by these sandwich degree graduates in the country. There was an insinuation that the quality of knowledge and skills acquired by these set of distance learning graduates are substandard to the ones acquired by residential graduates in conventional universities.

Managerial Formulation of Problem

As a follow up to this insinuation, Salami, et al (2004, 2006) decided to carry out a research among the employers of these sandwich degree graduations in Oyo State and Lagos State. The problem was however stated as follows:

> *that there is no significant difference in the quality of knowledge and skills acquired by sandwich degree graduates and the one acquired by residential graduates from universities at the labour market .*

Managerial Question Relating to the Problem

The university authority and the sandwich degree board of studies believe and stated that:

i. the quality of knowledge and skills acquired by the sandwich degree graduates in all the established outreach centres were sufficient and adequate to enable the recipients to perform eraditably well in the labour market.

ii. the process of moderating all aspects of evaluation system of sandwich degree programmes in all the outreach centres established by the university of Ado-Ekiti authority is the same with the residential, hence justifying the quality of knowledge and skills acquired by the two sets of graduates which is sufficient to allow them perform well in the labour market.

Statistical formulation of questions

Two major questions were raised to answer the managerial problems:

i. is there any significant difference in the perceptions of the public and the private employers of labour about the quality of knowledge and skills acquired by the sandwich degree graduates in Lagos State that would enable them perform well in the labour market?

ii. do employers of labour (private and public) disagreed on the quality of knowledge and skills acquired by the sandwich degree graduates in Lagos State that would enable them perform creditably well in the labour market?

Statistical Analysis

Population

The population of the study was made up of all establishments and organizations both private and public in Lagos State who had at least one graduate of the sandwich programme. Out of this population, officers in the rank of chief executive from seven (7) establishments were randomly chosen. viz: Principals and Vice Principals of Secondary Schools, Bank Managers, Local Government Chairmen, Divisional Police Officers, Headmasters of Primary Schools, LGEA Secretaries, PPTESCOM Supervisors and Directors of Education in Lagos State.

A total of two hundred (200) respondents in the five sampled Local Government Areas were randomly selected for the study in the ratio of sixty (60) Principals and Vice Principals, five (5) Local Government Chairmen, ten (10) Bank Managers, five (5) LGEA Secretaries, ten (10) Post-Primary Teaching Service Commission, Directors and Supervisors, one hundred (100) Primary School Headmasters, ten (10) Nigeria Police Officers in the rank of DPO.

Sampling Technique

Random sampling technique was used to pick respondents from Secondary and Primary Schools, Banks and Directors and Supervisors from PPTESCOM. On the other hand, the five LGEA Secretaries, five DPOs in the five sampled Local Government Areas were deliberately picked for the study.

The Instrument

Salami et al (2004) 12 item self-developed questionnaire was used for the collection of relevant information from the respondents. It has two sections. Section A is on the bio-data of the respondents. Section B contains 12 statements based on the study. The instrument had earlier being face validated by experts while its reliability coefficient was found to be 0.72. All the 200 questionnaires administered were correctly filled and returned giving a 100% returns.

Data Collection and Analysis

The researchers employed some teachers in Lagos State to administer the instrument. This was after briefing the research assistants as they were called on what to do and how to do it. To further simplify the analysis, we integrated the 4 Likert scale into two, namely Agree and Disagree. The data collected was analyzed, using simple percentages.

Answer to Statistical Questions

Research Question 1

To What extent do employers of labour perceive the quality of knowledge and skills acquired by the sandwich degree graduates in Lagos State as adequate and sufficient to enable them perform well in the labour market?

Table 1: Perception of public and private employers of labour on the quality of knowledge and skills acquired by the sandwich degree graduates in Lagos State. Nigeria.

S/N		AGREE		DISAGREE	
			%		%
1.	Graduates of Sandwich Degree Programmes in Lagos State compared favourably well with regular graduates in my organization	142	71	58	29
2.	The quality of programmes run by the Universities Sandwich Degree Centres in Lagos State is standard	130	65	70	35
3.	The performances and productivity of Sandwich Degree Graduates in my organization have increased since their completion of the study.	130	65	70	35
4.	Dedication and commitment to duty of Sandwich Degree Graduates have increased tremendously in my organization.	126	63	74	37
5.	Sandwich Degree Graduates in my organization are more proficient now than before the programme.	122	61	78	39
6.	I will support more of my staff members to undertake the Sandwich Degree Programme because of its quality.	128	64	72	36

Source: **Fieldwork work 2005**

From table 1 above, we concluded that employers of sandwich degree graduates in Lagos State accepted the fact that knowledge and skills acquired by these set of graduates were satisfactory and adequate. For instance 142 (71%) agreed that these set of graduates compared favourably with graduates who attended regular university programme. 130 or (65%) rated the quality of programmes run by the universities satellite campuses as standard, 130 (65%) confirmed that performances of sandwich graduates have increased after the completion of such programmes. While 126 or 63% confirmed the characteristics of dedication and commitment to duty of these graduates. Another 122 (61%) equally supported the proficiency of sandwich graduates. Finally, 128 respondents or 64% of the employers reached agreed to release more of their employees to undertake sandwich degree programmes anytime they want to go. Many of the employers interviewed confirmed that the programmes have given more opportunity to many who couldn t have gotten the chance to further their

academic studies. Through the sandwich programmes, many highly skilled manpower have been produced.

Research Question 2

Do employers of labour (private and public) disagree on the quality of knowledge and skills acquired by the Sandwich degree Graduates in Lagos State that will enable them perform creditably well in the labour market?

Table 2: Perceptions of the public and the private employers of labour on the quality of knowledge and skills acquired by the sandwich degree graduates in Lagos State.

S/N		AGREE		DISAGREE	
			%		%
7.	The performance and productivity of Sandwich Degree graduates in my organization have not changed from what they used to be.	84	42	116	58
8.	The quality of knowledge exhibited by Sandwich Degree graduates in my organization since the completion of the course is below expectation.	70	35	130	65
9.	The attitudes and commitment to duty of Sandwich Degree graduates in my organization have not changed from the previous behavior.	76	34	124	62
10.	Sandwich Degree graduates in my organization are not better in terms of qualitative knowledge than before programme.	90	45	110	55
11.	There is no noticeable change in the skill of Sandwich Degree graduates in my organization.	66	33	134	67
12.	The quality of programme run by University Degree Centres in Lagos State is below standard.	72	35	128	64

Source: Fieldwork work 2005

Table 2 above showed the responses of the employers of sandwich degree graduates in Lagos State. Reactions to 6 statements put to the employers revealed their disagreement with the statements put to them about the quality of knowledge and skills acquired by these set of graduates.

For instance, 116 (58%) of the respondents disagreed with the statement that performances of sandwich degree graduates in their organization have not changed. In the same vein 130 (65%) of the employers reached disagreed with the statement that performances of sandwich degree graduates are below expectation. Employers disagreement proved that these set of graduates performed excellently well in the labour market. In addition, 124 (62%), 110 (55%), 134 (67%) and 128 (64%) respectively disagreed with statements 9, 10, 11 and 12 as shown on the table.

Discussion of Findings

This study has also confirmed the adequacy of the quality of knowledge and skills acquired by the sandwich degree graduates in Lagos State. Just like the recent findings of Salami et al (2004) in Oyo State, Employers of labour both in the private and the public organizations in Lagos State have rated sandwich degree graduates as efficient as their counterparts who attended regular university programmes in Nigeria.

Implication of the Findings

Another major implication of Lagos State findings is that it corroborated the outcome of Salami et al (2004) findings in Oyo State. And that it has also proved scientifically that products of university sandwich degree programmes are equally good academically and professionally.

Answer to Managerial Questions

Therefore, this answer confirms the belief and affirmation of the university of Ado-Ekiti authority and sandwich degree board of Studies as to the quality of knowledge and skills acquired from Distant Learning Centres as adequate.

Managerial Solution to the Problem

Conclusion

The study has found out from the employers and users of sandwich degree graduates that products of sandwich degree programmes are good academically and professionally competent. The results from Lagos State findings corroborated early study carried out in Oyo State by Salami et al (2004). Higher percentage given to each statement further proved the assertion that these sets of products are professionally trained. Based on the findings, it is once again recommended that sandwich degree programmed anywhere in the country should be invigorated to accommodate more individuals yearning for higher qualification while retaining their jobs and that sandwich degree centers should equally maintain the high quality of academic excellence they have attained.

UNIT 3 BASIC STATISTICAL VOCABULARIES, NOTATIONS AND SYMBOLS

CONTENTS

1.0 Introduction

This unit will introduce the language of statistics to you. This is because statistics, like any other subject or discipline, has its own technical language. It is therefore always good and beneficial for education managers who will continuously use statistics to acquaint themselves with the commonly used vocabularies, notations and symbols of statistical language before the actual treatment of the subject matter. The unit begins with the commonly used terminologies or vocabularies.

2.0 Objectives

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

i. acquaint themselves with the commonly used terminologies, notations and symbols in statistics.
ii. differentiate between vocabularies, notations and symbol in statistics.
iii. identify the relationship between statistical symbol, notations and vocabularies.
iv. use the symbol to solve statistical equations and problems in the next units

3.0 Main content

3.1 Statistical Vocabularies

Vocabularies	Definition
Variable	A property which can assume any number.
Ordinary value	A variable on which scores or values are ordinarily recorded as number.
Categorized variable	Nominal variable; a variable on which positions or scores are not ranked.
Dichotomy	A categorical variable with only two categories
Score	Any position on a numerical variable.
Discrete variable	A variable generated by a counting process e.g. whole number
Continuous variable	A variable that can take on any value over a range of feasible value; measure data, can be whole number or fractions.
Interval variable	Variable that have equal interval along their scales of measurement
Ratio variable	Variable in which ratio of scores can be regarded as equal.
Data	Collected information, qualitative or quantitative
Raw Data	Data obtained directly from primary sources of information and which remained untreated, unprocessed and not yet manipulated
Distribution	The arrangement of a set numbers classified according to some property
Frequency Distribution	A table in which observed values of a variable are grouped or classified according to their numerical magnitude
Class frequency	Number of observations falling within the confines of a particular class.
Lower class limit	The lower possible value that can be assigned to a given class.
Upper class limit	The highest possible value that can be assigned to a given data.
Class interval	The difference between the lower and the upper limit of a given data.
Parameter	A descriptive measure for a population
Mean	Arithmetic average
Mode	The most commonly occurring value; the value of the variable that has the greatest frequency.
Median	The middle value of a set of figures.
Range	The difference between the smallest and the highest value in a distribution.

Standard Deviation	Root mean square deviation; a measure that describes the spread of a set of scores from the mean.
Kurtosis	Extent of peakedness of a distribution
Geometric mean	The antilog of the sum of the logarithms of the values of the factor comprising a group divided by the number of factor
Harmonic mean	The reciprocal of the arithmetic mean of the reciprocal of a set of values
Normal Distribution	A symmetrical distribution having its mean, mode and median equal one in which frequencies of the variable extend equally both to the left and the right of the mode.
Confidence Level	Degree of confidence for a given interval estimate
Parametric test	This is a test, which its efficacy rests on whether the variable being studied is at least approximately normally distributed.
Non parametric test	These are tests developed without references to the distribution of variables.
Hypothesis Testing	This is a method for testing whether a parameter equals a certain specified value.
Statistical Significance	Statistical prove of the difference between two means or variables or that the variation occur by chance
Significant Level	The probability that a statistics would be at least as large as an observed value by chance alone, if the hull hypothesis is true
Level one tailed significance	Probability of finding a certain result in one tail of a sampling distribution.
Level two tailed significance	Probability of finding a certain result in either of the two tails of a sampling distributions
Degree of freedom	This refers to the number of ways in which any set of scores is free to vary. It is also the number of restrictions placed on the set of scores.
Population	Population is the collection of universe to be studied .It may be finite, like the name of fall females in Nigeria .or infinite like the sand. it may also be real or hypothetical
Sample	A sample is a sub-set of a population
Target population	Population for which results are required
Bias	The difference between the value obtained by a researcher and the time value of the event.
Type I error	This is when a true null hypothesis is rejected instead of accepting it.
Type II error	This is when a false null hypothesis is accepted instead of being rejected.
Correlation	Is the statistical method used for establishing the extent of relationship or association between two or more scores.

3.2 Symbols or Formulae

	Symbols		*Definition*
1.	$\frac{\sum_{i=1}^{n}}{n}$ or $\frac{\sum x}{n}$		Arithmetic mean of ungrouped Data
2.	$\frac{\sum_{i=1}^{k} f_1 x_1}{n}$ or $\frac{\sum fx}{n}$		Arithmetic mean of grouped Data
3.	$G = \sqrt[r]{x_1 x_2 x_3 \ldots x_n}$		Geometric mean
4.	$H = \frac{N}{\sum \frac{1}{x}}$		Harmonic mean
5.	$M_q = \sqrt{\frac{\sum f x^2}{n}}$		Quadratic mean
6.	$a + \frac{(f - f_a)(b - a)}{(f - f_a) + (f - f_b)}$ or $l_o + \frac{f_o - f_1}{2f_o - f_1 - f_2}$		Mode
7.	$\frac{1}{2}(n + 1)^{th}$		Median of ungrouped Data
8.	$a + \frac{b - a}{f}(\frac{1}{2}n - Fa)$		Median of grouped Data
9.	$\frac{1}{n}\sum \lvert x - \bar{x} \rvert$		Mean Deviation
10.	$\frac{(x - \bar{x})^2}{2}$ OR $\frac{\sum x^2}{n} - \frac{(\sum x)^2}{n}$		Variance
11.	$\sqrt{Variance}$	(a)	Standard Deviation
12.	$\sqrt{\frac{(\sum x_2)}{n} - \frac{(\sum x)^2}{n}}$	(b)	Standard Deviation
13.	$\sqrt{\frac{\sum f(x - \bar{x})^2}{n}}$	(c)	Standard Deviation of grouped Data

14. $$r = \frac{\sum N XY - \bar{X} \quad Y}{\sqrt{[N\sum X^2 \quad (\sum X)^2] \; [N \sum^2 \quad (Y)^2]}}$$ Correlation coefficient using Raw Score method

15. $$r = \frac{\sum \quad \bar{x} \quad \bar{y}}{\sqrt{\sum (x - (x) \;) (Y \sqrt{\sum Y - \bar{y}}}}$$ $$)^2$$

OR $$r = \frac{\sum \bar{X}\bar{Y}}{\sqrt{\sum \quad \sum}}$$ $$x^2 \quad Y^2$$

Correlation coefficient using Deviation method

16. $$p = 1 \quad \frac{\sum}{N (N^2 \quad 1)} \; 6 D^2$$ Spearman Rank order correlation coefficient

17. $$X^2 = \frac{\sum}{E} \; (0 \quad E)^2$$ Chi-square coefficient for ungrouped Data or for Goodness of fit test

18. $$E (RC) = \frac{fr \times fc}{N}$$ Chi-square for obtaining expected frequencies of the cells OR for Test of Independence

19. $$\frac{\bar{X} \quad \bar{X}}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}}}$$ t or z statistical computation

20. $Y = a + bx$ Regression equation with least square method

21. $Y = a + bx_1 + b_2 x_2 + b_3 x_3 \; b_n x_n$ for multiple Regression Analysis

22. $$a = \frac{\sum \; \sum \quad \sum \quad \sum X}{(Y X \sum) \quad (X \sum Y) \quad (\quad)}$$ $$N X^2 \; (x)^2$$ For calculating the value of a using raw score method

23. $$b = \frac{\sum \quad \sum \quad \sum}{N XY \sum \; (X) \sum \; (Y)}$$ $$N X^2 \quad (X^2)$$ For calculating the value of b using raw score method

24. $Y = C + M$ Regression equation using Deviation method

25. $$M = \frac{\bar{x} \quad \bar{y}}{(x \quad x) \; (Y \quad)}$$ $$(x \quad)^2$$ For calculating the value of M in equation 23

26. $C = Y - M$ For calculating the value of c in equation 23

27. $$L_1 + \frac{(N_2 \quad Fa) \times c}{}$$ Alternative median formula for a grouped Data

fm

Self Assessment Exercise 1

Define the following terms:

i Range

ii Geometric Mean

iii Mode

iv Degree of Freedom

v Type 1 Error

Notations

X	Name of a random variable
F	Frequency of Occurrence
N	Number of observations
Fa	Cumulative frequency before the modal class
Fb	Cumulative frequency after the modal class
n	Sample size in the number of observations
Σ	Summation sign
$\bar{x}$	Arithmetic mean of a sample size
	Arithmetic means of a population
SD	Standard deviation of a sample size
L_1	Lower limit of class interval
L_2	Upper limit of a class interval
X_o	Assumed mean
df	Degree of freedom
H_o	Null hypothesis sign
H_1	Alternative Hypothesis sign
Md	Median
Mode	Mode
Fm	Frequency of a Median class
Q_1	First quartile

Q_3	Third quartile
n!	Factorial n
npr	number of permutation of n objects taken r at a time
$\binom{n}{r}$	Number of combination of r objects which may be taken from n objects.
O	Observed frequencies
E	Expect frequencies
R	Name of a discrete random variable
R^2	coefficient of determination
A, B, E	events.
A∩B	Intersection of the event A, B
A∪B	Union of events A, B.
A^c	Event complimentary to A
a, b, b_1, b_2	Estimates of regression coefficients ,$\propto$,β φ
$\propto$	alpha risk: probability of Type I Error
β	Beta risk: Probability of Type II Error
C	number of class interval
MD	Mean Deviation
Var.	Variance
X^2	Chi-square
E (RC)	expected frequencies of the roll and column cells
t-test	student t-distribution
Z-test	standardized normal score
SDx	Standard Error
t-cal	t-test calculated
t-crit	Critical or table value of t
cor (X,Y)	Covariance of X and Y
N (μ^2)σ	normal distribution with mean μ, variance^2 σ
S^2	estimated variance
F	Fisher s F distribution
T_+, T-	Wilcoxon signed Rank statistics

T	Wilcoxon rank sum statistics
U	Manu-Whitney statistic
t (n-1, 0.05)	95% point of t distribution with n-1 degree of freedom
$\hat{Y}$ $\hat{y}$	predicted values of Y, y in a linear model
β_1	regression coefficient of Y on X_1
λ	parameter of the exponential distribution or parameter of the poison distribution
$\hat{\lambda}$	maximum likelihood estimate of parameter
σ	Standard deviation of a random variable
σ^2	Variance of a random variable or variance of a population
Φ (b)	cumulative distribution function of normal distribution
p (E)	probability of an event E

4.0 Conclusion

In this unit, we have jointly examined the three major languages of statistics. Among these languages are statistical vocabularies, statistical symbols and statistical notations. Mastery of all these languages by the education managers become essential ingredient for the treatment of statistical problems which they will come across later in this course.

5.0 Summary

This unit has introduced to the students

* common terminologies usually found in statistics
* statistical symbols and formulae usually employed by statisticians
* notations that would assist the education managers in understanding the principles of statistics.

6.0 Tutor Marked Assignment

1. Give the definition of the followings

Raw Data, mode, standard deviation, Harmonic mean and Degree of freedom.

7.0 References/ Further Readings

Adewoye, S. O. (2004): Basic Statistics for Engineering, Economics and Management. Lagos: Olukayode Ojo Commercial Enterprises.

Hildebrand, D.K. & Lyman, O. (1998). Statistical thinking for managers (4th Ed.) California: Duxbury Press.

Levin, R.I. &Rubin, D.S. (1997). Statistics for Management (7th Ed.) New Jersey: Prentice Hall International.

Monga, G.S. (2007). Mathematics and Statistics for Economics (Second Revised Edition) New Delhi. VIKAS Publishing House PVT Ltd.

Salami, K.A. (1999). Descriptive Statistics for Beginners. Oyo Odumatt Press and Publishers.

Salami, K.A. (2001). Introduction to Basic Statistics in Management. In Adeyanju, A. (Ed.) Introduction to Educational Management. Oyo Green Light Press and Publishers.

Salami, K. A. (2006). Statistical Model and Projections in Educational Management. Lagos. Master Printers International.

UNIT 4 FREQUENCY DISTRIBUTION

CONTENTS

1.0 Introduction

This unit will introduce you to the technique of how to use frequency distribution and graph of various kinds to handle and interpret large masses of data which ordinarily couldn t have been possible through mere observations. With this technique, the large masses of data could be summarized in such a way as to reveal the important characteristics of the data at a glance. In addition, the unit will enable students to derive the essential features of the data with little effort.

Therefore, a brief outline of what you will find in this unit include how to arrange raw data in ascending and descending order, class interval, limit boundaries; mark and class mid point. The Histogram, frequency polygon and ogive are well treated. Types of frequency curves and Lorenz Curve of concentration are equally treated.

2.0 Objectives

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

1. arrange large number of raw scores in ascending and descending order.
2. identify various characteristics attributed to frequency distribution
3. construct histogram, frequency polygon, ogive and Lorenz Curves of concentration.
4. appreciate the functions of frequency distribution in statistics.

3.0 Main content

3.1 Raw data

These are data collected from primary source but have not been numerically organized, processed or manipulated.

Example of raw scores is given below.

Table 4.1: Raw scores of 30 students in a test.

25	30	30	30	25	25	50	50	50	55
60	60	60	60	60	65	65	65	65	40
45	45	40	40	45	48	48	50	50	30

3.2 Frequency Distribution

A frequency Distribution is a table in which the values of variables are classified according to size. Let the symbol for observations be X and the symbol for frequency distribution of scores be f as follows.

Example : Table 4.2 Frequency distribution of 30 raw scores

X	f	cf
25	3	3
30	4	7
40	3	10
45	3	13
48	2	15
50	5	20
55	1	2
60	5	2
65	4	30

We can also demonstrate this distribution with tally. The above frequency distribution is represented with tally below.

Table 4.3 Frequency Distribution with tally

X	Tally	f
25	III	3
30	IIII	4
40	III	3
45	III	3
48	II	2
50	IIIII	5
55	I	1
60	IIIII	5
65	IIII	4
		30

3.3 Arrangement of Data.

This is an arrangement of raw numerical data either in ascending or descending order of magnitude. Example of raw scores of 50 students in a test is given below.

Table 4.4 Raw scores of student in a test.

37	20	05	74	36	65	58	44
64	61	58	40	00	61	80	33
48	55	30	62	50	39	36	44
45	78	59	45	24	53	80	42
32	33	28	63	42	44	49	09
50	28	46	45	56	32	59	23
52	50						

Table 4.5: These raw scores are arranged in an ascending order as follows

00	05	09	18	20	23	24	28
28	30	32	32	33	33	36	36
37	39	40	42	42	44	44	44
45	45	45	46	48	49	50	50
50	52	53	55	56	57	58	58
59	61	61	62	63	64	65	74
78	80						

When summarizing large masses of raw data like the one we have in table 4.5, it is often useful to distribute the data into classes or categories and to determine the number of individuals belonging to each class, called class frequency. A tabular arrangement of data by classes together with the corresponding class frequency is also called frequency distribution. The scores in table 4.5 are re-arranged and put in frequency distribution using a class interval of 5.

Table 4.6 Arrangement of frequency Distribution using class interval of 5

Class interval f cf

0-4	1	1
5-9	2	3
10-14	-	3
15-19	1	4
20-24	3	7
25-29	2	9
30-34	5	14
35-39	4	18
40-44	6	24
45-49	6	30
50-54	5	35
55-59	6	41
60-64	5	46
65-69	1	47
70-74	1	48
75-79	1	49
80-84	1	50
	50	

3.4 Class interval and class limit

3.3.1 Class interval is the gap or number of digits between two figures. The interval may be three (3), five (5) or ten (10) as the case may be. Let us take 40-44 in table 4.6 as an example. With these two figures, the interval between them is 5.

3.3.2 Class limit refer to the lower limit and upper limit of two figures. With figures 40-44, the lower limit of the first figure is 39.5 while the upper limit is 44.5

3.5 Class Boundaries, Mark and Mid-Point.

3.5.1 Class Boundary and Mark

Theoretically, the class boundaries in a class interval between 40- 44 will include all measurements between 39.5 and 44.5. The 39.5 is the lower class mark while 44.5 is the upper class mark.

3.5.2 Class Mid-Point

The class midpoints is obtained by arranging scores (raw or processed) in ascending order and pick the middle one in ungrouped data. The class mid point in 41, 42, 43, 44, 45, is 43. In a grouped data like what we have in table 4.5, it is 2 for 0-4 group. e.g. 0-4 gives 0, 1,2,3,4. Here the class mid point is 2.

3.6 The Histogram

A histogram is a set of adjoining vertical bars whose areas are proportional to the frequencies represented by the bars. A histogram is drawn by taking the class interval on the X-axis and the frequencies on the Y- axis. When the class intervals are of equal width, the height of a bar corresponds to the frequency in that class. That is the class interval with the highest concentration of observations. This class is usually called the modal class.

The scale of axes are normally taken in such a way that the length of the Y- axis is about $1^{1/2}$times that of the X axis. This is to present a decent looking graph. Scores of 50 students in a test is presented in a histogram below.

Table 4.7: Frequency Distribution for a grouped data

Class interval	***Midpoint(x)***	***f***
25-30	27.5	3
30-35	32.5	4
35-40	37.5	4
40-45	42.5	5
45-50	47.5	15
50-55	52.5	7
55-60	57.5	3
60-65	62.5	4
65-70	67.5	2
70-75	72.5	1
75-80	82.5	1
		Total=50

Note that the highest frequency here is 15

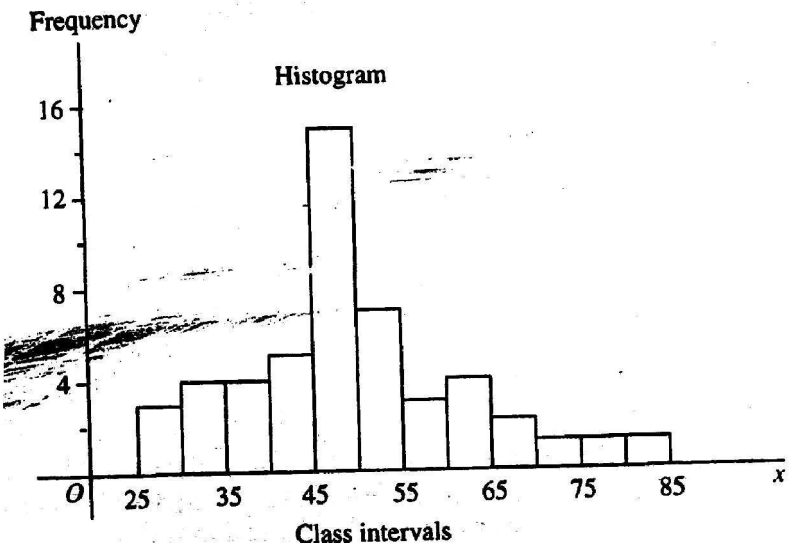


Fig. 4.1: A Histogram

3.7 The Frequency Polygon

The frequency polygon is the most frequently encountered graphic device in statistics. It is easy to construct and simple to interpret. The frequency polygon is a line chart plotted in the same way as the histogram. On the X-axis the class mid-Point are taken and frequency along the y-axis are represented by point, which are joined by straight lines to give us the frequency polygon. The first end point is joined to the X-axis to a midpoint showing zero frequency just before the first class interval and the last end joined to the one after the last class interval as shown in Fig 4.2

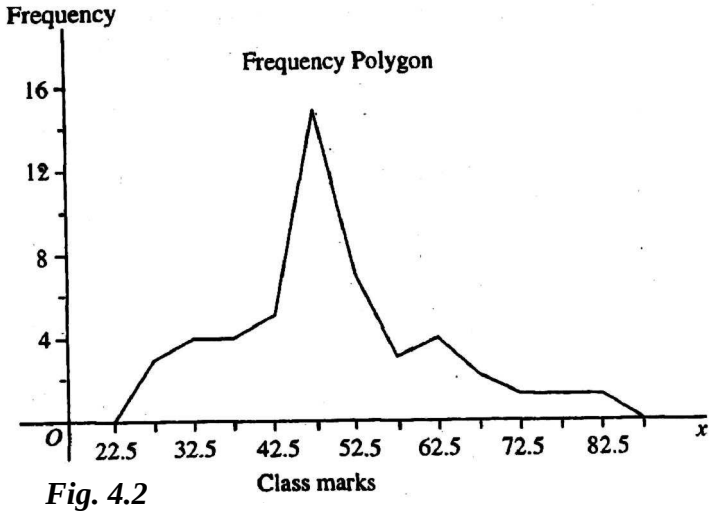


Fig. 4.2

The first midpoint 22.5 corresponds to zero frequency and the last i.e. 87.5 has zero frequency

3.8 Frequency Curves

A limiting form of the histogram or frequency polygon is the frequency curve. A smooth curve instead of broken lines joining the points corresponding to the frequency gives us the frequency curve of the data. It is a continuous curve and provides a frequency values for every value of x. The curve needs not necessarily pass through each and every point.

Fig. 4.3: A Frequency Curve

3.9 Relative Frequency Distribution

In a frequency distribution, if the frequency in each class interval is converted into a proportion by dividing by the total frequency, we shall get a series of proportions called relative frequencies. A distribution presented with relative frequencies rather than actual frequencies is called a relative frequency distribution.

The sum of all relative frequencies in a distribution is 1. The scores of 50 students in a test are shown below with its relative frequencies.

Table 4.7 Relative frequency distribution

Class interval	*Frequency x*	*Relative frequency*	*Explanation*
25-35	7	0.14	7/50 = 0.14
35-45	9	0.18	9/50 = 0.18
45-55	22	0.44	22/50 = 0.44
55-65	7	0.14	7/50 =0.14
65-75	3	0.06	3/50 = 0.06
75-65	2	0.04	2/50 = 0.04
	50	1.00	

Comment

The concept of relative frequencies is useful in sampling theory. It can also be used to compare two frequency distributions with unequal total frequency, though with the same series of class intervals. The example below illustrates this.

Class interval	f_1	f_2	*Rel.freq.*f_1	*Rel.Freq.*f_2
10-20	5	12	0.20	0.12
20-30	10	24	0.40	0.24
30-40	6	30	0.24	0.30
40-50	3	1	0.12	0.19
50-60	1	15	0.04	0.15
	25	100	1.00	1.00

Although, the first frequency is 5, while the second frequency is 10, but with the same series of class interval. The results of the two relative frequencies become 1.

Self Assessment Exercise 1

Arrange the following scores of 18 students in Economics test in (i) Ascending order (ii) Descending order and (iii) form frequency distribution with the data

33, 17, 26, 38, 40, 60, 75, 80, 17,

15, 12, 09, 24, 28, 30, 44, 60, 40.

3.10 Lorenz Curve of Concentration

The Lorenz curve is concerned with the distribution of any characteristics among the items possessing the same characteristics. The extent of concentration of the characteristics in a few items can be examined by the use of the Lorenz cu rve. An education manager can use Lorenz curve to illustrate the disparity between the rich people and the majority poor in a country or within educational system. Let us

consider the incomes received by some individuals, their numbers arranged according to class intervals. The classes are arranged in ascending order of income size and the cumulative totals of income and the number of individuals.

Table 4.9: Income Distribution Frequency

Income Range N	No of person (thousand) f	Total income (m) y	Cum Fr. Total Cum f	Cum total Of y. Cum y	% of Cum f	% Of Cum y
Over 5000	50	790	50	790	5	52.7
2000-5000	120	300	170	1090	17	72.7
1000-2000	200	220	370	1310	37	87.3
500-1000	180	100	550	1410	55	94.1
200-500	250	60	800	1470	80	98.0
100-200	200	30	1000	1500	100	100.0

The Line of Equal Distribution

The line of equal distribution in Lorenz curve says that if all groups of people had been receiving comparable proportion income e.g. 25 percent of people getting 25 percent of income and so on, then, the graph would be a straight line joining the points (0,0) (25,25) (100,100).

This will be the line of equal distribution

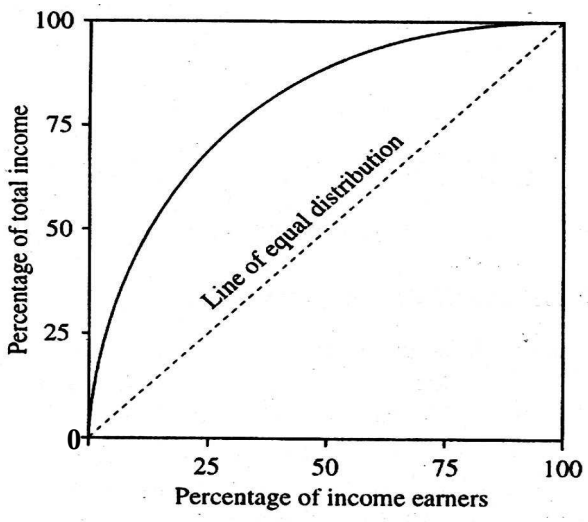


Fig. 4.4 The Lorenz curve

The degree to which a Lorenz curve deviates from th e line of equal distribution is a measure of the inequality of distribution of income. The farther the curve moves

away from this line the greater is the inequality. The degree of this inequality at any stage is indicated by the distance from the equal distribution line of the curve.

4.0 Conclusion

This unit has introduced you to many concepts in frequency distribution. Students have learnt how to use tally to construct frequency distribution and how to re-arrange raw score in ascending and descending order. The concepts of histogram, frequency polygon and frequency curve as well as their constructions were equally learnt in this unit

5.0 Summary

This unit has exposed students to:

- several concepts in frequency distribution; such as
- raw data
- arrangement of data
- class interval, limit, boundary, mark and mid-point
- histogram, frequency polygon and frequency curves & Lorenz Curve
- relative frequency.

6.0 Tutor Marked Assignment

1. Giving the scores of 30 students in Social Studies test

30	33	49	26	32	51	52	10	46	45
25	15	47	21	40	24	24	27	43	17
29	50	36	48	37	34	14	41	55	59

Use class interval of 5 and starting with the least score in the distribution, form frequency distribution table for the scores.

2. Compute the relative frequency distribution for the following data.

Class interval	***f***
20-30	3
30-40	5
40-50	6
50-60	7
60-70	10
70-80	09
80-90	8
90-100	3
	50

3. Write short note on the following:

i. class boundary and class mark

ii. class interval

iii. histogram

7.0 References/Further Readings

Adewoye, S. O. (2004): Basic Statistics for Engineering, Economics and Management, Lagos: Olukayode Ojo Commercial Enterprises.

Hildebrand, D.K. & Lyman, O. (1998). Statistical thinking for managers (4th Ed.) California: Duxbury Press.

Levin, R.I. &Rubin, D.S. (1997). Statistics for Management (7th Ed.) New Jersey: Prentice Hall International.

Monga, G.S. (2007). Mathematics and Statistics for Economics (Second Revised Edition) New Delhi. VIKAS Publishing House PVT Ltd.

Salami, K.A. (1999). Descriptive Statistics for Beginners. Oyo Odumatt Press and Publishers.

Salami, K.A. (2001). Introduction to Basic Statistics in Management. In Adeyanju, A. (Ed.) Introduction to Educational Management. Oyo Green Light Press and Publishers.

Unit 5 MEASURES OF CENTRAL TENDENCY

CONTENTS

1.0 Introduction

This unit will introduce you to the concept of average, Calculation of arithmetic mean is and the uses of such averages. Most times, education managers are confronted with the problem of finding the average mark of a class, the average wage bill, the average cost of a building and usually the average expenses per week, or month. In addition, the manager may want to know the amount of wage bill that is frequently or constantly paid. It is on this premise that the knowledge of mean, mode and median and other averages becomes imperative and essential to the education managers and essential for proper school administration.

2.0 Objectives

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

i. calculate mean, mode and median from groups of data

ii. identify the properties of a arithmetic mean

iii. understand the uses of averages in education

iv. identify the merits and demerits of averages

3.0 Main Contents

3.1 Definition of Average

Quantitative data arranged in the form of frequency distribution generally exhibit a common characteristic, that is, they have the tendency to concentrate at certain values, usually somewhere near the centre of the distribution. The tendency of observations to cluster near the central part of the distribution is known as Central Tendency and can be measured statistically.

An average therefore is a precise yet a simple expression representing a series of divergent individual values, in other words, it is the consolidated essence of a complex distribution.

3.2 The Arithmetic Mean

The arithmetic mean is obtained by dividing the sum of values of observations by the number of observations. It is usually denoted by $\bar{x}$. It is also the best known and most commonly used form of average.

3.2.1 Simple Arithmetic Mean

Example 5.1 For the observation 5, 4, 9, 12, 10, the arithmetic mean is obtained as follows.

$$\bar{x} = \frac{5+4+9+12+10}{5} = \frac{40}{5} = 8$$

x is a variable having n values

In formula form, it is written as:

$$\bar{x} = \frac{x_1 + x_2 \quad x_n}{n} = \frac{\sum_{i=1} x_1}{n} \text{ or } \frac{\sum x}{n}$$

3.3 Arithmetic Mean of Grouped Data

When observations repeat themselves, we may reduce the cumbersomeness involve by considering their frequencies.

Example 5.2 For the observation 5 5 5 4 4 3 3 3 3, the Arithmetic mean may be written as

$$\bar{X} = \frac{3_1 \text{ x } 5 + 2 \text{ x } 4 + 4 \text{ x } 3}{3 + 2 + 4} = \frac{35}{9} = 3.9$$

OR Add all the observations together and divide by the number of variables such as

$$\frac{5+5+5+4+4+3+3+3+3}{9} = \frac{35}{9} = 3.9$$

If x is a variable having values

$x_1, x_2 \quad x_n$

and occurring with frequencies

$f_1, f_2 \quad f_k$

then, its arithmetic mean for that grouped data may be written

$$\frac{\sum_{i=1}^{k} f_1 x_1}{n} \quad \text{or} \quad \frac{\sum fx}{n}$$

Example 5.3 The frequencies of the wages of 20 workers is given as shown in the table below:

Table 5.1

Wages X	**Number of Workers** f	fx (N)
5	2	10
7	4	28
9	5	45
11	6	66
13	2	26
15	1	15
Total	**20**	**190**

It means 20 workers received a total of 190 naira in the form of wages. Therefore, the Arithmetic mean of this wage is

$$\bar{x} = \frac{190}{20} = N9.50$$

The mean can be computed with the help of the following formula

$$\bar{x} = \frac{\Sigma fx}{n} = \frac{190}{20} = ₦9.50$$

Arithmetic mean of wages is ₦9.50

3.4 Properties of Arithmetic Mean

1. The Arithmetic mean is affected by all the items in the series, which is good in so far as it makes $\bar{x}$ a very representative average. It is not a good representative if there is a large deviation from the central value.
2. It is easy to calculate and is capable of algebraic manipulations.
3. It is determinate
4. It is a typical representative value of all items in the data.
5. Its value may be substituted for the value of each item without changing the total:

e.g. $\Sigma fx = \Sigma f\bar{x} = n\bar{x}$

It is very important to education managers to note that the algebraic sum of the deviations of a set of numbers from the arithmetic mean is zero i.e.

$$\Sigma f(x - \bar{x}) = \Sigma fx - \bar{x}\Sigma f = n\bar{x} - n\bar{x} = 0$$

And if there are no frequencies, we can write

$$\Sigma x = \Sigma\bar{x} = n\bar{x}$$

$$\Sigma(x - \bar{x}) = n - n = 0$$

Example 5.4 The Arithmetic mean of 3,10,15 and 8 is = 9 that is, $3+10+15+8 = \frac{36}{4} = 9$

Therefore (3 9) + (10 9) + (15 9) + (8 9) = 0

By working with the formula

$$\Sigma f(x - \bar{x}) = \Sigma fx - \bar{x}\Sigma f = n\bar{x} - n\bar{x} = 0$$

$$36 - 4x9 = 39 - 9x4 = 4x9 - 4x9 = 0$$

$$36 \quad 36 = 36 \quad 36 = 36 \quad 36 = 0$$

3.5 Merits of Arithmetic Mean

The Arithmetic mean is the most widely used measure of central tendency because:

i. its definition is clear and precise, it corresponds to the centre of gravity of the observations.

ii. it is simple to understand and easy to compute

iii. it makes use of each and every item in the data.

iv. it has a determinate value and it is rigidly defined.

v. it can be subjected to further algebraic treatment and advanced statistical theory.

vi. it can be found even if only the total values is known and the individual values are not known.

vii. it provides a good standard of comparison since extreme values cancel each other out when the number of observations is large.

3.6 Demerits of Arithmetic Mean

i. it could be influenced by unrepresentative values. For example, the mean of 49, 50, 51 is 50. But the mean of 1,50,99 is also 50. In such cases, the representative character of the mean is lost.

ii. it gives greater importance to larger and less importance to smaller values. It has an upward bias.

iii. it cannot be calculated if one or more items in the data are missing.

iv. it cannot be located by inspection (like the mode & median).

v. it is capable of cancelling facts and such may lead to distorted conclusions.

3.7 Geometric Mean (GM)

The geometric mean (GM) is the nth root of the product of n values. In formula form it is written.

$$G = \sqrt[n]{x_1, x_2, x_3, \quad x_n}$$

The G.M of 2, 4, 8 is the cube root of their product

Solution

$$G \sqrt[3]{2.4.8} = \sqrt[3]{64} = 4$$

The G.M of 1, 2, 3 is the cube root of their product

$$G = \sqrt[3]{1.2.3} = \sqrt[3]{6} = 1.82$$

The G.M of 2,4,4,8 is the cube root of their product

$$G = \sqrt[3]{2.4.4.8} = \sqrt[3]{256} = 4$$

Logarithms may also be used in the calculation of G.M. If the frequencies of x_1, x_2 xk are respresentative of f_1, f_2 f_k (Σf =n) then,

$$\text{Log } G = \frac{1}{n}[f_1 \log x_1 + f_2 \log x_2 + \quad + fk \log fk] = \frac{\Sigma f \log x}{n}$$

$$G = \text{Antilog } \frac{1}{n} \Sigma f \log x$$

But if there are no frequencies, G = (x_1, x_2 x_n) and $\log G = \frac{1}{n} \sum \log x$

Example 5.5 For value of x without frequencies, we compute as follows:

x	log.x
4.5	0.6532
250.5	2.3958
12.0	1.0792
119.5	2.0774
30.0	1.4771
42.0	1.6232
75.0	1.8751
35.4	1.5490
	12.7330 = $\sum log\ x$

G = Antilog $\frac{1}{n}\sum log\ x$ = Antilog $\frac{1}{n}$ x 12.7330 = Antilog 1.5916 = 39.05

Example 5.6 Value of x with frequencies

Class interval	**f**	**Mid-point x**	**log x.**	**f log x.**
2 4	20	3	0.4771	9.542
4 6	40	5	0.6990	27.960
6 8	30	7	0.8451	25.353
8 10	10	9	0.9542	9.542
	$\sum f$ = 100			$\sum f \log\ x$ = 72.397

G = Antilog $\frac{1}{n}\sum f$ log x

= Antilog $\frac{1}{100}$ (72.397) = Antilog 0.72397

G = 5.296

3.7.1 Merit and Uses of Geometric Mean

Most of the properties and merits of G.M resemble those of Arithmetic mean (A.M).

i. The G.M. takes into account all the items in the data and condenses them into one representative value.

ii. It has a downward bias. That is, it gives more weight to smaller values than to larger values.

iii. It is determinate. That is, for the same data there cannot be two geometric means.

iv. It balances the ratios of the values on either side of the data. It is ideally suited to average, rates of change such as index numbers and ratios between measures and percentages.

v. It is amenable to algebraic manipulations like the Arithmetic mean (AM)

3.7.2 Demerits of Geometric Mean

i. It is difficult to use and compute

ii. It is determined for positive values and cannot be used for negative values or zero. A zero will convert the whole product into zero.

3.8 The Harmonic Mean (HM)

The Harmonic mean is the reciprocal of the arithmetic mean of the reciprocals of a series of observations.

In other words, the total number of items of a variable in a series, divided by the total of the reciprocals of the items gives the H.M.

The formula for computing H.M. is

$$H = \frac{N}{\sum \frac{1}{x}}$$

and if we have a frequency distribution

$$H = \frac{n}{\sum \frac{f}{x}}$$

Example 5.7 The Harmonic mean of 2, 3 5 is

The Arithmetic mean of the reciprocal is 1/3 (½ + 1/3 + 1/5)

$$= 1/3 \frac{(10 + 15 + 6)}{30}$$

$$= 1/3 \left(\frac{31}{30}\right)$$

$$= \frac{31}{90}$$

The Harmonic mean therefore, is the reciprocal of the Arithmetic mean of the given number

$$H = \frac{90}{31}$$

$$H = 2.9$$

Example 5.8 The Harmonic mean of 3, 5, 6 is

H = 1/3 (1/3 + 1/5 + 1/6)

= 1/3 $\left(\frac{21}{30}\right)$

= $\left(\frac{21}{90}\right)$

The Harmonic mean therefore is the reciprocal of the Arithmetic mean which is

$$H = \frac{90}{21}$$

H = 4.3 approximately

Example 5.9 The Harmonic mean of 0.5 and 0.25 is

$$H = \frac{2}{1 + \frac{1}{0.5}} = \frac{2}{\frac{1}{0.25} + 4}$$

= 1/3 or 0.33

For the value of x with frequencies, the Harmonic mean could be calculated as shown in the example below:

Example 5.10

Class interval	Mid point x	$\frac{1}{x}$	f	fx $\frac{1}{x}$
2 4	3	0.333	20	6.67
4 6	5	0.200	40	8.00
6 8	7	0.143	30	4.29
8 10	9	0.111	10	1.11
			$\sum x$ = 100	$\sum fx\frac{1}{x}$ = 20.07

$$H = \frac{n}{\sum \frac{n}{x}} = \frac{100}{20.07} = 4.98$$

3.8.1 Merit and Uses of Harmonic Mean

i. Like any computed average such as A.M. and G.M, the H.M. also takes into account all the observations in the data.

ii. It gives more weight to smaller items i.e. it has a downward bias. It is also of great utility when weights are attached to smaller items.

iii. It measures rates of change and can be adapted to problems involving time and certain type of ratios and rates.

iv. It is amenable to algebraic manipulations.

3.8.2 Demerits of Harmonic Mean

i. It is difficult to compute when the number of items is large.

ii. Some people, particularly the users may find it difficult to understand because it is not in common use as an average.

iii. It assigns too much weight to the smaller items and thus has limited scope.

iv. It is interesting to note that its value is always less than that of the Geometric Mean (GM), which is usually smaller than the Arithmetic Mean (AM). The relative positions of AM, GM, and HM may be symbolically represented as follows.

$$\underline{x >} \; \underline{G > H}$$

Note that the three means are identical, if all the items in a given data have the same value.

3.9 The Quadratic Mean or The Root Mean Square

It is the square root of the mean of the square values of x in a series of k observations.

The formula is

$$Mq = \sqrt{\frac{x_1^2 + x_2^2 + \quad + x_k^2}{k}} = \sqrt{\frac{\sum x^2}{k}}$$

The formula for weighted quadratic mean with weights or frequencies f_1, f_2 fk for the values x_1, x_2 x_k respectively where $\sum f = n$ is given as

$$Mq = \sqrt{\frac{\sum f x^2}{n}}$$

Example 5.11 The Quadratic mean of 3, 2, 4, 5, 1, 7 is

$$Mq = \sqrt{\frac{1}{6}(3^2 + 2^2 + 4^2 + 5^2 + 1^2 + 7^2)}$$

$$= \frac{\sqrt{104}}{6}$$

$$= 4.16$$

Example 5.12 The Quadratic mean of 2 3 4 6 8 is

$$Mq = \sqrt{\frac{1}{5}(2^2 + 3^2 + 4^2 + 6^2 + 8^2)}$$

$$= \frac{\sqrt{104}}{5}$$

$$Mq = 5.08$$

3.10 The Mode

The mode is the value which occurs most often in data. It is the value around which there is the greatest degree of concentration.

The mode means norm or fashion . The mode is thus a typical measure of central tendency in as much as it is the most probable value in the series.

Example 5.13 The most in this series 3, 7, 3, 3, 5, 3, 1, 2, is 3. In this series, 3 appears four times, so it becomes the mode.

Self Assessment Exercise 1

(i) Identify the number of Modes in this set of figures.

1,7,6,2,3,4,5,4,5,4,5,

(ii) Find the Median of this set of ungrouped figures.

4,2,3,3,7,5,9

(iii) Find the Geometric mean of the following

3,4,6,8

3.10.1 Bimodal

A set of scores or figures could contain more than one mode. For example, with a set of 1, 2, 3, 4, 4, 4, 5, 5, 5, 6, 7. The modes are 4 and 5 because 4 and 5 appeared three times each than other figures. These two figures (4 and 5) are called bimodal .

3.10.2 Modal Class

The class in which the mode falls is called the modal class. It corresponds to the highest frequency.

Example 5.14 The group scores of 34 statistics students in a test is given as follows:

Table 5.2: The Group Score

Class interval	*f.*	*cf.*
52 56	03	03
56 60	6	09
60 64	10	19
64 68	04	23
68 72	08	31
72 76	02	33
76 80	01	34
	34	

The modal class of this grouped data is between 60 and 64 because the number of students whose scores fall between 60 and 64 are ten (10).

3.10.3: Mode by Interpolation

However the formula for finding the modal class mark of a grouped data is given as

Method 1

$$\text{Mode} = a + \frac{(f \quad fa)(b \quad a)}{(f \quad fa) + (f \quad fb)}$$

where

a = first number of a modal class
fa = frequency number before the modal class
fb = frequency number after the modal class
f = frequency numbers of the modal class

Example 5.15 From table 5.2 Find the modal class mark

Solution

From the data, a = 60, b = 64, f = 10, fa = 6, fb = 4

Therefore, proper substitution will give us

$$\frac{60 + (10 \quad 6)(64 \quad 60)}{(10 \quad 6) + (10 \quad 4)}$$

$$\frac{60 + 4 \times 4}{4 + 6}$$

$$\frac{60 + 1.6}{10}$$

$$60 + 1.6$$

Modal class mark = 61.6

Method 2

$$\text{Mode} = l_o + \frac{f_o \quad f_1}{2f_o \; f_1 \quad f_2}$$

where

l_o = the lower limit of the modal class
c = the size of the class interval
f_o = the frequency of the modal class
f_1 = the frequency of the class immediately below the modal class

Note that a distribution with one mode is called unimodal; with two modes, it is called bimodal. With many modes we have a multimodal distribution. Using the same figures in table 5.2 We have the modal class as between 60 and 64 therefore

$l_o = 60$, $c = 4$, $f_o = 10$, $f_1 = 6$, $f_2 = 04$

with good substitution we have

$$m_o = l_o + \frac{f_o \quad f_1}{2f_o \quad f_1 \quad f_2} \times c = 60 + \left[\frac{10 \quad 6}{2x10 \quad 6 \quad 4}\right]4$$

$$= 60 + \left[\frac{4}{20 \quad 6 \quad 4}\right]4$$

$$= 60 + \frac{4 \times 4}{10}$$

$$= 60 + \frac{16}{10}$$

Model class mark = 61.6

3.10.4 Merits and uses of Mode

i. It is simple to define and compute
ii. It is a popular average in the sense that it is the one that most people use without being aware of it.
iii. Extreme values have no effect on the mode and it can be calculated when complete data are not available.
iv. It is the most typical in the sense that it denotes the most probable value in the series.

3.10.4 Demerits of Mode

i. In the case of bimodal and multimodal distributions, it is not possible to pinpoint any one value as the mode.
ii. It is not rigidly defined and thus cannot be called an ideal average.
iii. It is often indeterminate when the distribution is highly irregular.
iv. It is not based on all the observations in the data and hence is not an ideal measure of central tendency. It lays two much emphasis on the modal group and thus may not be fully representative of the whole series.

v. It is not easily amenable to algebraic manipulations.

3.11 The Median

The Median divides the data into two equal parts so that 50% of the items lie on either side of it. However to get the Median, a set of figures must be arranged in either ascending or descending order.

3.11.1 Median Formula for Ungrouped Data

When the number of n observations is odd and the observations are arranged in ascending order, the Median is the ½ $(n + 1)^{th}$ observations or simply the middle value.

Example 5.16 Find the Median of this set of ungrouped Data

44, 40, 79, 42, 51, 59, 71, 44, 60, 65, 45

Solution. Re-arrange the numbers in ascending order like this.

40, 42, 44, 44, 45, 51, 59, 60, 65, 71, 79

Then apply the formula = ½ (11+1) = $\frac{12}{2}$ = 6

Therefore, the 6^{th} number which is 51 is the Median

Example 5.17 Find the Median of 4, 2, 3, 5, 3, 6, 9, 7

Solution Re-arrange the numbers in ascending order like this

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

Then apply the formula ½ (8 + 1) = $\frac{9}{2}$ = 4.5

Therefore, with even numbers, the Median lies between figures 4 and 5 so that 4.5 or 4 ½ becomes the Median .

3.11.2 Median Formula for Grouped Data

The median of grouped data can be obtained through the use of two formulae

1. Med = $a + \frac{b \quad a}{f}(\frac{1}{2} n \quad Fa)$

where

a = first number of the Median class

b = second number of the Median class

f = frequency number of the Median class

n = total number of variables involved

Fa = cumulative frequency number before f

2. Med = $L_1 + \frac{(N/2 = F_1) \times C}{}$

fm

where

L_1 = lower class mark
N = total number of variables involved
fm = frequency number of the Median class
F_1 = cumulative frequency number before (fm)
c = number of class interval

Example 5.18 Find the Median of the set of scores of 34 statistics students in a test given below

Class interval	***f.***	***cf.***
52 56	3	3
56 60	6	9
60 64	10	19
64 68	4	23
68 72	8	31
72 76	2	33
76 80	1	34
	$\sum f$ = 34	

Solution

1. Using the first formula we get

$$60 + \frac{64 \quad 60}{10}(17 \quad 9)$$

$$60 + 3.2$$

Med = 63.2

Solution

2. Using the second formula, we get

$$\text{Med.} = 59.5 + \frac{(34/2 \quad 12)}{10} \times C$$

$$= 59.5 + \frac{(17 \quad 12)}{10} \times C$$

$$\frac{+ (5) \times 5}{10}$$

$$\frac{+ 25}{10}$$

$$59.5 + 2.5$$

Med. 62

Note: Median Class is 60- 64, a =60, b=64, f =10 and fa =9

Students are to note that the differences occurred as a result of fractions used in the second formula.

3.11.3 Merit and Uses of the Median

i. it can be easily understood and its computation is simple.
ii. it can be computed even for imcomplete data. It is concerned only with a few central observations.
iii. it balances the number of items in a distribution
iv. it is useful in describing scores, ratios and grades
v. it is useful in the case of skewed distribution like those of income and prices.
vi. it can be used for qualitative data.
vii. in the case of open end classes, the Median can be calculated but the mean cannot
viii. it can be easily determined graphically.

3.11.4 Demerits of the Median

i. The median is not easily capable of algebraic manipulations. As such it is not much used in advanced studies.
ii. The empirical formula for the Median based on interpolation may not always give correct results.
iii. it may ignore significant extreme values.
iv. weighting cannot be used in the case of the Median.
v. The scope of operations is thus narrowed.
vi. it can not be computed as exactly as the mean.

4.0 Conclusion

In this unit, we have discussed almost all aspects of measures of central Tendency. Particularly, the unit highlighted the three major commonly used averages viz: Arithmetic mean, mode and median and the other three, Geometric mean, Harmonic mean and Quadratic mean which are of benefit to education managers. The properties of Arithmetic mean as well as the Merits and Demerits of all the means were equally learned.

5.0 Summary

This unit has exposed students to:

i. the definition of Average
ii. the computation of arithmetic mean, mode and median
iii. the benefit of knowing how to compute and use Geometric mean, Harmonic mean and Quadratic mean.
iv. the advantages, disadvantages and uses of all these means.

6.0 Tutor Marked Assignment

1. Compute the Arithmetic mean wage for the following workers

Wages	***no of workers***
x	**f**
10	5
17	7
19	11
22	8
33	5
55	3
	39

2. Find the modal class mark for the following data.

Weight of student		***frequency***
x		**f**
90	100	10
100	110	37
110	120	65
120	130	80
130	140	51
140	150	35
150	160	18
160	170	04
		300

7.0 References/Further Readings

Adewoye, S. O. (2004): Basic Statistics for Engineering, Economics and Management, Lagos: Olukayode Ojo Commercial Enterprises.

Akinwumiju, J. A. (1988): Basic Statistical Researching and Interpretation in Educational Research, Ibadan: University of Ibadan.

Amos, J. R. et al (1965): Statistical Concepts (A Basic Program). New York: Harper & Row Publishing.

Cooke, D., Craven, A. H. and Clark, G. M. (1990): Basic Statistical Computing 2nd ed, Arnold, London (CCC).

Hildebrand, D.K. & Lyman, O. (1998). Statistical thinking for managers (4th Ed.) California: Duxbury Press.

Levin, R.I. &Rubin, D.S. (1997). Statistics for Management (7th Ed.) New Jersey: Prentice Hall International.

Monga, G.S. (2007). Mathematics and Statistics for Economics (Second Revised Edition) New Delhi. VIKAS Publishing House PVT Ltd.

Salami, K.A. (1999). Descriptive Statistics for Beginners. Oyo Odumatt Press and Publishers.

Salami, K.A. (2001). Introduction to Basic Statistics in Management. In Adeyanju, A. (Ed.) Introduction to Educational Management. Oyo Green Light Press and Publishers.

Unit 6 MEASURES OF VARIABILITY

CONTENTS

1.0 Introduction

In unit five, measures of central tendency as averages of the first order was extensively treated. In this unit, measures of variability as the averages of the second order will be treated. More importantly, a measure of dispersion gives a more idea about the extent of lack of uniformity in the sizes and qualities of the items in a series. It helps us to know the degree of uniformity and consistency in the series.

In a nutshell, you will be introduced to the computation and uses of Range, Quartile deviation, Mean deviation, Standard deviation, Variance and Kurtosis. The relevance of all these in educational management is fully highlighted.

2.0 Objectives

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

i. define the concept of Range
ii. compute Quartile, Mean and Standard Deviations
iii. compute Variance
iv. identify various Kurtoses from graphs

3.0 Main Content

3.1 Dispersion

Dispersion is designed to measure variation, that is, the extent to which individual item in a group are dispersed or distributed over the whole range of the independent variable.

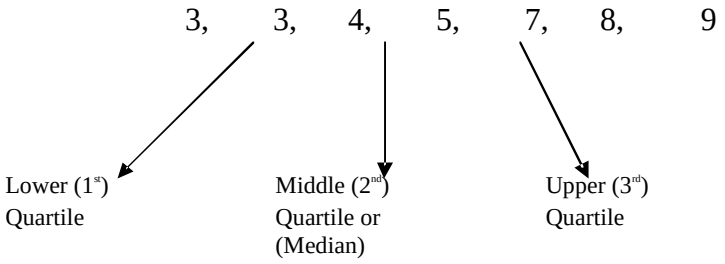
3.1.2 Skewness

Measures of skewness provides a measure of the symmetry in the distribution.

3.2 Quartiles

The identification of first, second and third quartiles can best be demonstrated with example as shown below.

Example 6.1: The set 7,4,5,3,3,9,8can be size-ordered and the quartiles identified as follows.



Therefore, identification of the quartiles from an ordered set of data is just as the median can be identified from the value of the $\frac{n+1^{th}}{2}$ item

Q_1 is the value of the $\frac{n+1^{th}}{4}$ item

Q_3 is the value of the $\frac{3(n+1)^{th}}{4}$ item

Example 6.2: Find Q_1 and Q_3 from the company s expenditure given below.

Expenditure (₦)	Number of companies	Upper bound	cf	cf%
Less than 500	210	500	210	16.1
500 to 1000	184	1000	394	30.0
1000 to 1500	232	1500	626	48.0
1500 to 2000	348	2000	974	74.0
2000 to 2500	177	2500	1151	88.3
2500 to 3000	83	3000	1234	94.7
3000 to 3500	48	3500	1282	98.4
3500 to 4000	12	4000	1294	99.3
4000 and over	09	5000	1303	100

$Q_1 = \frac{n+1^{th}}{}$ item $= 1304 = 326$

$$Q_3 = \frac{3(n+1)}{4}^{th} \text{ item} = \frac{3(1304)}{4} = 978$$

3.3 Range

The range (R) of a set of numbers is the difference between the maximum and minimum values. It indicates the limits within which the values fall.

Example 6.3 The range of the series, 5,15,17,20,24 is R = 24-5=19

Relative Range or the coefficient of range is defined by the ratio

$$RR = \frac{Xmax \quad Xmin}{Xmax + Xmin} = \frac{24-5}{24+5} = 0.66$$

3.3.1 Merit of Range

(i) It is easy to understand.
(ii) Its computation is simple.

3.3.2 Limitation of Range

(i) Since it is based on two extreme values in the entire distribution, the range may be considerably changed if either of the extreme cases happens to drop out.

(ii) It does not take into account the entire data.

(iii) It can not be computed when the distribution has open-end cases.

3.4 Quartile Deviation

Quartile deviation is not a measure of dispersion in the sense that it does not show the scatter around an average, but only a distance on scale. Consequently, quartile deviation is regarded as a measure of partition.

Students are to note that aside for the fact that its computation is simple and it is easy to understand, a quartile deviation does not satisfy any other test of a good measure of variation.

$$Q.D = \frac{Q_3 \quad Q_1}{2}$$

The relative measure is given by the

$$Q.D = \frac{Q_3 \quad Q_1}{Q_3 + Q_1}$$

Example 6.4 Given the following data, $Q_1 = 5$, $Q_3 = 10$, find Q.D and Coefficient of Quartile deviation (Q.D).

$$Q.D. = \frac{10-5}{} = \frac{5}{} = 2.5$$

$$\text{Coefficient of Q.D} = \frac{10-5}{10+5} = \frac{5}{15} = \frac{1}{3}$$

Example 6.5: Find the Q.D. and coefficient of Q.D. for the following data.

Earnings in (₦)	f	cf
10-20	42	42
20-30	25	67
30-40	58	125
40-50	42	167

Key: f =frequency, cf = Cumulative frequency

$$Q_1 = 10 + \frac{42 - 0}{42} \times 10 = 20.00$$

$$Q_3 = 40 + \frac{126 - 125}{42} \times 10 = 40.23$$

$$\text{Q.D} = \frac{40.23 - 20.00}{2} = 20.23 = 10.115$$

$$\text{Coefficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{20.23}{60.23} = 0.33$$

3.5 Mean Deviation

This measures the average or mean of the sum of all deviations of every items in the distribution from a central value; usually the mean or median. The larger the differences between the items, the larger the mean deviation will be. If all the items are identical, the mean deviation will be zero.

The mean deviation is only concerned with measuring the extent of the deviation about the mean, the signs preceding the deviations are ignored in the calculations and hence this measure of dispersion is of little use in practice.

3.5.1 Steps necessary for the computation of mean deviation

(i) Calculate the median of the array.

(ii) Record the deviation /d/ of each in a group from the median.

(iii) Multiply each deviation with their respective frequencies.

(iv) Add these deviations /d/ while you ignore the signs.

(v) /d/ should be divided by the total number of items (n).

Or use the following formulas:

$$M.D = \frac{\sum / x - \bar{x} /}{n}$$

for grouped data $M.D = \frac{\sum f / x - \bar{x} /}{n}$

Coefficient of $M.D = \frac{M.D}{\bar{x}}$

Mean deviation from the median = $\frac{\sum / x - me /}{n}$

and for a grouped data $\frac{\sum f / x - me /}{n}$

Students are to note that if the M.D is calculated with reference to the median, the coefficient of M.D becomes M.D/Mean.

Example 6.6 Find the M.D from the mean of the series, 5,7,10,12,6

x	$/ x - \bar{x} /$
5	3
7	1
10	2
12	4
6	2
$\sum x = 40$	12

$n = 5$

$\bar{x} = 8$

Deviation = $\sum$ / x $\bar{x}$=12

$M.D = \frac{\sum / x / \bar{x} /}{N} = \frac{12}{5} = 2.4$

Coefficient of $M.D = \frac{M.D}{\bar{x}} = \frac{2.4}{8} = 0.3$

Example 6.7: From the Median and the co-efficient of M.D. is calculated from the data given below:

Age (Years)	No of persons f	Mid-value (x) x	/x me/	f/ x me/

1 5	7	3	17	119
6 10	10	8	12	120
11 15	16	13	7	112
16 20	32	18	2	64
21 25	24	23	3	72
26 30	18	28	8	144
31 35	10	33	13	130
36 40	5	38	18	90
41 45	1	43	23	23
	123			$\Sigma f/x$ me/=874

In unit five, we gave the formula for Median, which is 2nd quartile (Q_2) as

$$l_o + \frac{\frac{n}{2} - F}{fo} \times c$$

Applying this formula to the grouped data above gives

$$\text{Median} = 16 + \frac{\frac{123}{2} - 23}{32} \times 5$$

Me = 16 + 5.937

Me = 21.937

Therefore, we have the M.D from the Median as

$$= \frac{\Sigma f/x - me/}{n} = \frac{874}{123} = 7.11$$

Coefficient of M.D = $\frac{\text{M.D}}{\text{Median}} = \frac{7.11}{21.937} = 0.32$

3.6 Variance

The variance of a set of observations $x_1, x_2 \quad x_n$ is the average of the squared deviations from their means. In formula form, it is denoted by

$$\frac{1}{N} \sum_{l=1}^{n} (x_1 \quad \bar{x})^2$$

3.6.1 How to Calculate Variance

The first step in working out the variance is to calculate the square of the deviation of each number. The variance is the mean of the squared deviation.

Example 6.8 Find the variance of 61, 52, 55, 58, 54

x	$(x - \bar{x})$	$(x - \bar{x})^2$
61	+ 5	25
52	- 4	16
55	- 1	01
58	+ 2	04
54	- 2	04
$\Sigma x = 280$		50

n = 5

$\bar{x} = 56$

Variance = $\frac{50}{5} = 10$

Example 6.9 Find the variance of 2,4,6,8,10

x	$(x - \bar{x})$	$(x - \bar{x})^2$
2	- 4	16
4	- 2	04
6	0	00
8	+ 2	04
10	+ 4	14
$\Sigma x = 30$		40

n = 5

$\bar{x} = 6$

Variance = $\frac{40}{5} = 8$

3.6.2 Another formula used for computing variance is denoted by

$$\text{Var.} = \frac{(\Sigma X^2)}{n} - \frac{(\Sigma X)^2}{n}$$

Example 6.10: Use this formula to find the variance using the data in example 6.8

x	x^2
61	3721
52	2704

55	3025
58	3364
54	2916
280	15,730

n = 5

var. = $\frac{15730}{5} - \left(\frac{280}{5}\right)^2$

= 3146 3136

= 10

The same answer with what was obtained in example 6.8

Self Assessment Exercise 1

Use this variance formula: Var = $\frac{(\Sigma^2)}{n} - \frac{(x\Sigma^2}{n}$

to compute the variance for 2,4,6,8,10

3.7 Standard Deviation

Standard deviation is the most satisfactory and most widely used measure of dispersion.

3.7.1 Procedure for Calculation

(i) Write down the squares of the deviation from the mean. This will of course, be positive.

(ii) Calculate the mean of these squares of deviations.

(iii) The standard deviation is the square root of (ii) above.

Students are to note that the standard deviation is always measured from the mean and not from the median or mode.

Standard deviation is the positive square root of the variance and the formula is given as

SD. = $\sqrt{\frac{1}{N}\sum_{i=1}(x - x)^2}$ Or $\sqrt{\frac{©fd^2}{©f}}$

Example 6.11: Find the SD. for Data in examples 6.8 and 6.9 using this formula

Variance in 6.8 is 10

Standard deviation is $\sqrt{10}$ = 3.16

Variance in 6.9 is 8

Standard deviation is $\sqrt{8}$ = 2.83

3.7.2 Standard Deviation of Ungrouped Data

The two commonly used methods for calculating the standard deviation of ungrouped data are:

$$\sqrt{Variance} \qquad (1)$$

$$\sqrt{\frac{(\Sigma x_2)}{n} - \left(\frac{\Sigma}{n}\right)^2} \qquad (2)$$

$$\sqrt{\frac{\Sigma fd^2}{\Sigma f}} \qquad (3)$$

3.7.3 Essence of Standard Deviation

A small standard deviation tells us that the observations cluster closely around their means, while a large standard deviation says that the observations are much more scattered. The standard deviation is a very commonly used measure of variability. It does use all the observations, and because the variance can be studied mathematically, it is possible to develop theoretical result or develop managerial problems and result involving the standard deviation.

3.7.4 Standard Deviation from Grouped Data

The formula for calculating standard deviation of a grouped data is given as:

$$SD = \sqrt{\frac{\Sigma f (x - \bar{x})^2}{N}}$$

3.7.5 How to Calculate Standard Deviation with this Formula

(a) **Using Deviation method**

i. Find the deviation of each score from the mean $(x - \bar{x})$

ii. Square each deviation $(x - \bar{x})^2$

iii. Multiply each squared deviation by the corresponding frequency $f(x - \bar{x})^2$

iv. Add up the results in (iii) $\Sigma f (x - \bar{x})^2$

v. Divide the sum obtained in (iv) by n

vi. Find the square root of the result of (v)

Example 6.12: Calculated the standard deviation for the data below using deviation/fraction method.

(a) **Deviation method**

x	f	fx	$(x - \bar{x})$	$(x - \bar{x})^2$	$f(x - \bar{x})^2$
10	5	50	+2.70	7.29	36.5
9	6	54	+1.70	2.89	17.34
8	9	72	+0.70	0.49	4.41
7	10	70	+0.30	0.09	0.90
6	3	18	+1.30	1.69	5.07
5	3	15	2.30	5.29	- 15.87
4	2	8	- 3.30	10.89	- 21.80
3	1	3	- 4.30	18.49	- 18.49
2	1	2	- 5.30	28.09	- 28.09
	40	292			148.47

n = 40

$\bar{x} = \frac{292}{40} = 7.30$

Substituting the figures into the formula, we have

$$SD = \sqrt{\frac{\Sigma f (x - \bar{x})^2}{N}} = \sqrt{\frac{148.47}{40}}$$

$$= \sqrt{3.71}$$

SD = 1.93

(b) **Raw Score Method**

Another method with which we can calculate the standard deviation of a set of figures is known as the Raw Score method with its formula as:

$$SD = \frac{1}{N} \sqrt{\Sigma f x^2 - \frac{(\Sigma f x)^2}{N}}$$

Note that the steps involved in applying this formula are as follows:

i. Multiply each score by its frequency

ii. sum the result in (1) to give ($\Sigma f x$)

iii. Square the result of (ii) to obtain ($\Sigma f x)^2$

iv. Square each score in the distribution to get x^2

v. Multiply the result of each squared score by its frequency to obtain fx^2

vi. Sum the result in (v) to get Σfx^2

vii. Multiply the result in (vi) by iv to obtain N Σfx^2

viii. Subtract the result of (iii) from the result of (vii)

ix. Find the square root of (viii)

x. Divide the result of (ix) by N to obtain the standard deviation

Example 6.10: compute the standard deviation for the same figures in Exercise 6.9 using raw score method.

x	f	fx	x^2	fx^2
10	5	50	100	500
9	6	54	81	486
8	9	72	64	576
7	10	70	49	490
6	3	18	06	108
5	3	15	25	75
4	2	8	16	32
3	1	3	09	09
2	1	2	04	04
	40	292	384	2280

$$SD = \frac{1}{N}\sqrt{N\Sigma fx_2 \quad (\Sigma fx)^2}$$

$$= \frac{1}{40}\sqrt{91200 \quad (292)^2}$$

$$= \frac{1}{40}\sqrt{91200)} \quad 85264$$

$$SD = \frac{1}{40}\sqrt{5936}$$

$$\frac{1}{40}(77.0454)$$

= 77.0454

SD = 1.93

3.7.6 Application of Measures of Variability

i. Measures of variability in general give insight into the performance of a given group of students. A small variability indicates small variation which implies that the students performances are markedly different.

ii. Information on measures of central tendency and measures of variability would enable us to determine whether our students have attained mastery in a unit of instruction. In this wise, a high mean and moderate standard deviation would indicate that the students have achieved mastery, while low mean and low standard deviation would indicate that mastery was not achieved in the unit of instruction.

Self Assessment Exercise 2

1. Compute the mean, modal class mark, median and mean deviation from the data below:

Class interval		**f**
15	19	4
20	24	16
25	29	12
30	34	24
35	39	14
40	44	12
45	49	10
50	54	08
55	59	06
60	64	04
		110

2. If a university in Nigeria has six different labour unions with the following number of officers 25, 30, 18, 27, 28, 22 respectively find the mean, variance and standard deviation of the distribution.

3.8 Kurtosis

Kurtosis is the measure of peaked ness of a distribution. It shows the degree of convexity of a frequency curve.

If the normal curve is taken as the standard, symmetrical, bell shaped curve, kurtosis gives a measure of departure from the normal convexity of a distribution

3.8.1 Types of Kurtosis

- The normal curve is mesokurtic. It is of intermediate peakedness
- The flat-topped curve, broader than the normal, is called platykurtic
- The slender, highly peaked curve is called leptokurtic

3.8.2 Measures of Kurtosis

i. Moment coefficient of kurtosis: $\beta_2 = \frac{\mu_4}{\mu_2^2}$

Note. Instead of β_2, statisticians often use $Y_2 = \beta_2 - 3$ the outcome of which is positive for a leptokurtic distribution; negative for a platykurtic distribution and zero for the normal distribution.

ii. Percentage coefficient of kurtosis: $K = \frac{Q}{P_{90} - P_{10}}$

where $Q = \frac{1}{2}(Q_3 - Q_1)$ which is the same interquartile range.

Example 6.12: Find the skewness and kurtosis for the following distribution by the method of moment.

Number of Hours worked	Number of days
1 – 3	3
3 – 5	5
5 – 7	1
7 – 9	1
	10

To get the mean ($\bar{x}$ for this data, we need to find the midpoint.

No of Worked	Hours	Midpoint (x)	No. of days	$\sum fx$
1	3	2	3	6
3	5	4	5	20
5	7	6	1	6
7	9	8	1	8
				40

For the data the $\bar{x} = \frac{\sum fx}{n} = \frac{40}{10} = 4$

if we substitute d for x ; $\bar{x}$ then our required table becomes

x	f	d	d^2	fd^2	fd^3	fd^4
2	3	-2	4	12	-24	48
4	5	0	0	00	0	0
6	1	+2	4	4	+8	16
8	1	+4	16	16	+64	256
	10			32	48	320

Explanation

x - midpoint of the distribution

f - the frequency of the distribution

d - x $\bar{x}$

d^2 - d^2

fd^2 - d^2 x f

fd^3 - d x fd^2

fd^4 - d x fd^3

the analysis of the distribution is

$S^2 = \mu_2 = \frac{1}{10} \sum fd_2 = \frac{1}{10} \times 32 = \sqrt{3.2}$ therefore s = 1.79

$\mu_3 = \frac{1}{N} \sum fd \quad \frac{1}{10} \times 48 = 4.8$

$\mu_4 = \frac{1}{n} \quad \frac{1}{10} \times 320 =$

32.0

Skewness: $\gamma_1 = \beta_{23} = \frac{\mu_3^2}{\mu_2^3} = \frac{4.8^2}{3.2^3} = \frac{9}{128} = 0.0703$

Kurtosis: $\beta_1 \frac{\mu_4}{\mu_2^2} = \frac{32.0}{3.2^2} = \frac{32.0}{1024} = 3.125$

Alternatively: $Y_1 = \gamma_1 = \sqrt{\beta_1} = \sqrt{0.0703} = 0.265$

$Y_2 = \beta_2 - 3 = 0.125$

Note: The Skew ness is positive and the distribution is leptokurtic as $Y_2 > 0$

4.0 Conclusion

This unit has again introduced you to another concept of variability which is called measures of variability or dispersion. Going through this unit you were able to identify the range between two poles or figures, Quartile, Quartile deviation as well as mean deviation were equally treated for you. Particularly, you were acquainted with computation of variance and standard deviation and the relationship between the two variabilities was highlighted. A measure which is rarely used but very important, the skewness or kurtosis and how to measure the kurtosis was given special treatment in this unit.

5.0 Summary

In this unit, adequate treatment was given to:

i. range and its identification

ii. measurement of several deviations such as Quartile, Mean deviation, Variance and Standard Deviation.

iii. kurtosis which shows the level of skewness and peakedness of any distribution.

iv. the uses of all these variabilities in education.

v. how education managers can use these deviations for decision making.

6.0 Tutor Marked Assignment

1. Calculate the mean deviation from the following data.

Class Interval	f
0-10	18
10-20	16
20-30	15
30-40	12
40-50	10
50-60	5
60-70	2
70-80	2
	80

2. Use Raw Score method to compute the standard deviation for the following data.

x	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21
f	2 4 5 8 10 9 11 3 4 2 2

3. Discuss the merits and limitations of range.

7.0 References/ Further Readings

Adewoye, S. O. (2004): Basic Statistics for Engineering, Economics and Management, Lagos: Olukayode Ojo Commercial Enterprises.

Hildebrand, D.K. & Lyman, O. (1998). Statistical thinking for managers (4th Ed.) California: Duxbury Press.

Levin, R.I. &Rubin, D.S. (1997). Statistics for Management (7th Ed.) New Jersey: Prentice Hall International.

Monga, G.S. (2007). Mathematics and Statistics for Economics (Second Revised Edition) New Delhi. VIKAS Publishing House PVT Ltd.

Salami, K.A. (1999). Descriptive Statistics for Beginners. Oyo Odumatt Press and Publishers.

Salami, K.A. (2001). Introduction to Basic Statistics in Management. In Adeyanju, A. (Ed.) Introduction to Educational Management. Oyo Green Light Press and Publishers.

MODULE TWO

Unit 1 The Probability and Non- Probability Sampling.

Unit 2 Probability Theory and Distribution

Unit 3 Estimation

Unit 4 Testing of Hypothesis

CONTENTS

1.0 Introduction

The theory of sampling distribution basically establishes the basis for statistical estimation. A statistician takes a number of observations for the measurement of a certain phenomenon and repeat his observations a number of times. Every time, he depends on a sample of observations to be able to make some general statement about the result. This is because in practice, he cannot study all the conceivable cases. Attempting to observe all the cases may be senseless and perhaps impossible. For example, to find out the attitude of all senior secondary school students to mathematics in Nigeria will be time consuming, energy wasted and costly whereas a good sample randomly selected among the SSS 3 students in the country will equally give a fairly good idea about the attitude of the whole population of students.

The focus of this unit therefore, is to present to you how to select samples that will be representative of the population from which it is drawn.

2.0 Objectives

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

(i) identify concepts closely related to sampling distribution

(ii) draw samples from different large populations.

(iii) identify various random samplings in education.

(iv) differentiate between sample mean and population mean.

(v) appreciate how the mean of a large sample approximate to the population mean.

(vi) identify the differences between probability sampling and non-probability sampling

(vii) list the advantages and disadvantages of both sampling techniques

3.0 Main Content

3.1 The Basic Concepts

3.1.1 Population

Ordinarily, the term population refers to a group of people inhabiting a specified geographical location. This is the case when we talk of the population of Nigeria, the population of Oyo State, and so on. In research, the term is used in a more general sense to include all members or element, be they human beings, animals, trees, objects, events etc of a well defined group. Population defines the limits within which the research findings are applicable. In other words, the population is defined in such a way that the results of the investigation are generalizable unto it.

Population can be classified into two, They are target and accessible population. The target population is all the members of a specified group within which the investigation relates, while the accessible population is defined in terms of those elements in the group within the reach of the researcher. For instance, a target population may be defined to include J.S.S. students in Nigeria, while the accessible population may be J.S.S. students in Oyo State, if these are the students within the researcher s reach.

The factor, which determines the choice of population is the problem under investigation. The population should be such that it can provide the most authentic and dependable data necessary for solving the problem. Again, it should be such that the generalizations or conclusions from the study can validly apply to it. The broader the definition of the research population relative to the actual portion of the population involved in the study, the less valid will the generalizations from the study which apply to that population. In other words, generalizations apply more validly to the accessible population than they would apply to the target population. The external validity of any study (i.e. the generalizability of the study), is usually judged in terms of how validly the conclusions of the study apply to the population specified for such a study. So, when a researcher is specifying his research population, he is setting some standard against which his study will be judged.

3.1.2 Sample

For some studies, the group of items to which the study relates (i.e. the population) may be small enough to warrant the inclusion of all of them in the study. But a study may entail a large population, which cannot all be studied. That portion of the population that is studied is called a sample population. A sample is, therefore a smaller group of elements drawn through a definite procedure from a specified population. The elements making up this sample are those that are actually studied.

Based on the data obtained from the sample, generalization or inferences on the population is made. Drawing inference or generalization about the population beside on the data obtained from the sample is of primary concern in any scientific investigation. Knowledge of the sample is of little significance in itself if such knowledge cannot be extended to the population.

3.1.3 Variables: Quantitative and Qualitative

When a characteristic takes different values, it is called a variable. A variable that can be measured is called quantitative variable.

A variable that cannot be measured but can only be categorized as belonging to some others is called qualitative variable.

3.1.4 Random Variable

A random variable is a quantity whose value is determined by a chance experiment. It is also called a chance variable or stochastic variable. In technical terms, a random variable is defined as a measurable function determined on a probability space.

3.1.5 Finite and Infinite Population

Finite population has definite number of elements that are countable, while infinite population has an indefinite number of elements. The elements in infinite population are so many that we cannot determine or count them. The number of school in Oyo State may be cited as an example of finite population. On the other hand, the number of leaves in a forest may be regarded as an infinite population.

Self Assessment Exercise 1

i. Differentiate between population and sample

ii. Give examples of finite and infinite population

3.1.6 Parameter

This is the numerical characteristic of a population. For example, the population mean is N, the population standard deviation is d, and the population proportion is P and so on.

3.1.7 Random Sample

A random sample is the selection of each object from the population in such a manner that gives equal chance of being included in the random sample. Every object is as likely to be considered as any other.

3.2 Sampling Distribution

If we can compute a statistic for each sample size N drawn from a given population (such statistic as the mean, standard deviation etc) which varies from sample to sample, then wee have a distribution of the statistic which is called its sampling distribution.

3.2.1 Estimating the Sample Size

The size of a sample to be taken in a survey is very crucial and must be determined with high consideration. This is because such an exercise relates to the

time, cost and efficiency factors. We should note that a very large sample may result into a waste of resources and time, while a very small sample may not give enough information. Therefore, the sample size to be taken depends on:

(i) the desired precision and the limits of error to be allowed.

(ii) the properties of the population.

(iii) the kind of sampling to be used.

(iv) the purpose.

(v) the resources, time and labour available and the costs to be incurred.

3.2.2 Sampling Techniques

A sampling technique is a plan specifying how element will be drawn from the population. Sampling techniques may be categorizes into two major types viz: the probability sampling techniques and the non-probability sampling techniques.

Sampling with replacement is when each member of the population is free to be selected at all times and sampling without replacement implies that each member cannot be selected more than once.

3.3 Probability Sampling Techniques

Probability sampling techniques means that each observation in the population has an equal chance of being selected to become a part of the sample. In this category are: Simple random sampling; stratified sampling; cluster sampling; multistage sampling; systematic sampling and sequential sampling. Personal judgment has no role to play in probability sampling.

The precision of a probability sample increases with the size of the sample. Some of these sampling techniques will be discussed in this unit.

3.3.1 Simple Random Sampling

In this type of sampling, each of the population has equal and independent chance of being included in the sample. If in a population, there are 400 elements, the probability or chance of drawing each element is 1/400. For a population of size 3000, the chance of drawing each element is 1/3000. Having independent chance of being included implies that the chance of drawing an element does not depend on (or is not affected by) the drawing of another element.

Sample resulting from the application of this procedure are said to be unbiased and are therefore representative of the population, at least theoretically. On the other hand, sample drawn in such a way as precludes some of the elements from being included, or makes the inclusion of some elements more likely or more probable than others, are said to be biased and therefore not representative of the population from which they were drawn.

The random sampling is by far the easiest and simplest probability sampling technique, in terms of conceptualization and application. It does not necessarily require knowledge of the exact composition of the population, so long as we can reach all members of the population. Since no further classification of the population is necessary, researcher does not require a thorough knowledge of the population characteristics to be able to carry out this type of sampling. Therefore, the errors usually arising from improper classification (known as classification errors do not occur). Under this sampling plan, the sampling error is usually known and can be precisely determined.

3.3.2 Methods of Drawing Random Samples

Simple random sampling can be carried out through any of the following means:

(a) use of slip of paper

(b) use of table or random digit

(c) use of computer

(a) On each of these, the name or an identification mark of one member of the population is written. The slips are folded and put in a container. After thorough reshuffling, the researcher, not looking in to the container, dips his hand and picks one slip. He unfolds the slips, records the element it contains, fold it again and puts it back into the container. This process is repeated until he draws the required number of elements.

This process whereby, after each draw, the slips are put back into the container before subsequent draw is referred to as sampling with replacement. It ensures that each member of the population has an equal probability of being drawn; any element that has been drawn once is ignored whenever it is drawn on subsequent occasions. If

a slip that has been drawn is not put back into the container before the subsequent draws, the process is said to be sampling without replacement. In this case, the element will not have the same probability of being included.

(b) **Use of table or random digits:** A table of random digits is a continuous sequence of numbers not appearing in any particular order. No number in the sequence appears more often that the other. To use this table, all elements in the population are numbered. The researcher, having pre-determined the size of the sample he wishes to draw, enters the table at any point. Beginning from this point, the researcher decides to move upwards or sideways. The elements having those numbers drawn from the table now constitute the sample.

(c) **Use of the Computer:** the computer is instructed to print a series of numbers as many as the desired sample size. These elements whose numbers are so printed define the sample. The use of the computer for this purpose is particularly useful when the population size is large.

Self Assessment Exercise 2

i. Illustrate with examples, two methods through which we can draw samples from a population.

- **Random Sampling with Replacement**

If each selected unit is returned to the population after each draw, every item continues to have the same chance of being drawn. If the population is of a finite size N, this procedure may be employed.

Exercise 11.1

A population consists of the four numbers 2, 5, 9, 12. Let us consider all possible samples size of two which can be drawn with replacement from the population.

Solution

There are $4^2 = 16$ samples size of two which can be drawn with replacement (since any one of the four numbers on the first draw can be associated with any one of the four numbers on the second draw). These are:

(2,2)	(2,5)	(2,9)	(2,12)
(5,2)	(5,5)	(5,9)	(5,12)
(9,2)	(9,5)	(9,9)	(9,12)

(12,2) (12,5) (12,9) (12,12)

- **Random Sampling without Replacement**

If each selected unit is not returned back into the population and after each draw, each of the remaining items has an equal chance of being drawn. In such situation, we have sampling without replacement.

This method is utilized when the population size is infinitely large. Many researchers, however, always adopt the simple method of placing all the population in a container and pick the required number from the container for their study.

Exercise 11.2

Using the same four numbers 2,5,9,12, let us see the possible samples size of two which can be drawn without replacement from the population.

Solution

There are ${}^4C_2 = 6$ samples of size two which can be drawn without replacement (this mean that we draw one number and then another number different from the first) from the population.

(2,5), (2,9), (2,12)

(5,9), (5,12)

(9,12)

4C2 means 4 Combination 2

3.3.3 Stratified Random Sampling

Stratification is employed when a heterogeneous population can be subdivided into such strata as are nearly homogeneous themselves. In such situation, a prior knowledge of the population is required.

Stratified sampling amounts to taking a number of simple random samples from subpopulation which is the strata of the given population. Stratified sampling gives the most efficient results when the variability within each stratum is the least; that is, each stratum is more or less homogenous within itself. A good example of stratification in education is when we stratified schools into A, B, C or D. Another example is when the population of males in a country is stratified into the old, young and infants respectively. As education manager, it is administratively more convenient to deal with strata rather than a whole population.

3.3.4 Multistage Sampling

Multistage sampling is useful and appropriate when a survey is to be carried out over a wide area involving heavy travel expenses. But first and foremost:

(i) the population must be divided first into large groups or first stage units, one of the first stage units is first selected to be used in the second stage.

(ii) this is divided into smaller, second stage units and so on till we arrive at a reasonable sized group to be utilized for the selection of the ultimate sample.

3.3.5 Cluster Sampling

The use of a sampling unit which consists of a group or cluster of the elements in the population is known as cluster sampling. The selection of the sample is done in

two stages. First, certain groups or clusters are selected from the population. These are primary sampling units. From each of these clusters, elementary sampling units are drawn. From this follow alternative names for this method of sampling; namely, sub sampling or two stage sampling, or block or area sampling.

Cluster sampling permits groups of observations for easier coverage. Travel or other expenses may be significantly reduced and thus the cost per elementary unit becomes much less. We should note that the results of a cluster sample may not be as precise as those of a random sample but they can be made more precise by taking a larger sample size.

3.3.6 Systematic sampling

Systematic sampling represents a particular case of cluster sampling in which the sample is a single cluster. In this case, a sample is selected at a predetermined sampling interval. Thus, every 10^{th} unit in the population gives us a 10 per cent sample. In general, every nth unit may be selected. This method has an advantage over random sampling if neighboring observations resemble one another. A good example is when an education manager wants to interview university graduates and NCE (Nigeria Certificate in Education) teachers for teaching employment. A systematic selection of few from each group will be time and money saving if it is handled carefully.

3.3.7 Sequential Sampling

Sequential sampling as the name implies refers to a system of testing a small number of items or interviewing few people from many units or population and accepting or rejecting the whole lot depending on the decision taken with the help of the sample. It is possible to reach a positive decision on the whole after interviewing the first batch, but in case no decision is reached from the first sample, then, more samples are taken till a decision is possible. However, sequential sampling is used mostly in production unit to sample manufactured products. Nevertheless, the knowledge of it is equally good to know by the education managers who are expected to manage not only human beings but also material resources as well.

Non- Probability Sampling

3.3.8 Purposive or Judgment Sampling

Personal judgement can sometimes be more useful than a probability sample in pilot surveys or small scale surveys. Judgement sampling is used if the person in charge is well experienced and is also known to have good power of judgement. For example, when it comes to the construction of index number, the judgement of experts in the field are more useful than random method.

Another example of judgement sampling is when we want to find out about the performance of a particular principal of a school, and to do this, only those teachers who have a bachelor s degree with 5 years post qualification experience are selected for the exercise while other teachers are excluded. Such sampling becomes judgemental.

From these examples you will see that judgement sampling can be biased because there is no objective test to determine the validity of its results. Also its utility may not necessarily increase with the increase in sample size.

Another good example of judgement sampling is when we want to estimate per capita income of a country and the investigator decided to interview only the car owners. This selection will be seriously biased because car owners may generally be richer than other non car owners in such a country. Again such car owners will not be a good representative of the population. Therefore, using such judgement sampling will not give us the true picture of per capita income in that population.

3.3.9 Quota Sampling

Quota sampling does not permit probability method because each investigator is instructed to collect information from an assigned number or quota of individuals having specified characteristics. This type of sampling is used to ensure that specific elements will be included. Consider the population of school principals in Nigeria. In quota sampling, the investigator simply lists the number of principals from each state he wishes to include in the sample. If he wishes to use interview method, he goes to each state and interviews the desired number of principals from those he can easily reach.

This type of sampling allows the investigator to include any category of the population that is of particular interest to him. Doing that gives the investigator great score for bias. Although and perhaps in many cases, quota sampling has proved to be a very convenient and less costly method. And that this sampling is commonly used for market survey and public opinion polls on issues like winning an election etc.

3.3.10 Accidental or Convenience Sampling

In accidental or convenient sampling, only elements which the investigator can reach are included in the sample. The only determining factor is the investigator s convenience and economy in terms of time and money. The consideration is not whether these elements possess some specific characteristics or not. This is where convenience sampling differs from quota sampling. The television reporter who interviews any person he sees in a particular area is carrying out accidental sampling.

Apart from the convenience and the economy of time and money, this sampling method has no other advantage. This is because the resultant sample is totally biased since it does not represent or appear to be a representative of the population from which it is drawn. It will be very difficult to draw correct generalizations based on such sample.

3.3.11 Simulation

You will agree with me that human problems in real life are often very complex and may not be easily expressed in the form of comprehensible mathematical equations. In such a situation, simulation technique can be employed to arrive at the solution to such problems.

The technique of simulation consists of a series of organized trial-error experiments with the model of a system to enable one to derive certain results when the mathematical format of the model is not easy. Simulation helps in prediction and decision making under conditions of uncertainty to provide solutions to problems arising in a variety of business and managerial situations where values of variables are not available or not easily available or when mathematical equation could not help because of complications of various types.

Let us consider the example of a nagging partner in a couple. This situation relates to attitude or behavior arising from uncountable factors and which no known mathematical or statistical equation can represent. Although, Monte Carlo methods as well as systems simulation method have been put in place, non has actually succeeded in solving human real world problems.

3.4 Error

The concept of error should be taking with all seriousness when we engage in statistical measurements, hypothesis testing and decision making. This is because there is usually a possibility of errors. In fact, without mistakes or commitment of errors here and there in human life, there would be no development.

Therefore, we should know that some errors are systematic and are independent of the size of the sample. They may be related to defects in planning stage, procedure, methods of collecting data, faulty questionnaire design etc. The term commonly used for systematic error is bias.

The other type of errors of interest to us is random or sampling error which arises from differences between the results of sample observations and universe values. The samples may be drawn form the population using the same methods, yet differences may occur. The important thing to note is that the size of this error indicates the reliability or precision of the experiment. As the sample size increases, this error decreases and hence larger samples are considered more reliable than smaller samples. The true values of any measurement is thus associated with Bias (Systematic Error) + Random Error each of which may be positive and negative.

Self Assessment Exercise 1

Is the error in the following random or systematic? A selected number of 1000 people were asked in questionnaire whether the present politicians are corrupt, honest or otherwise. Of the 50 people that replied, a majority did not consider their honest. But the majority of the people are known to consider their corrupt.

3.5 Advantages of Sampling

A sample provides partial information about the population and yet, as we have seen from units eleven and twelve, we have to depend, for various reasons on the information available from the sample.

Among the advantages of sampling are:

i. **Practicability.** Often a complete census is impracticable. It way, therefore not be carried out at all. Selecting samples becomes the next option.

ii. **Flexibility.** Sampling has greater scope and flexibility in the matter of types of information to be obtained.

iii. **Accuracy.** More and greater accuracy is achieved with more efficient workers whoa re equally given intensive training particularly to a small number of workers who will perform less work. Often sampling methods do give accurate results than a complete survey.

iv. **Speed.** The collection and analysis of data can be done more quickly if the data are not excessive. Also time and energy are saved.

v. **Superfluity Avoided.** Often there are many items of the same type. A complete enumeration will consider each one of them. That will amount to waste of time and energy.

vi. **Reduced Cost.** Expenditure on a sample is likely to be very small as compared with that on complete enumeration as the case in population census.

4.0 Conclusion

In this unit, students have been exposed to some concepts which are germane to the understanding of probability sampling. For example, that the existence of a population is often required as an assumption so as to allow for the selection of samples which are central to probability sampling. Again, the assumed population can be finite or infinite and that in most cases, if not all, statistics deal with finite population. Furthermore, students were enlightened also that probability sampling concerns itself with the scientific method of selecting samples without bias. The differences in the concepts associated with probability sampling and non-probability sampling were equally highlighted

5.0 Summary

This unit has enlightened you on:

(i) many concepts germane to the understanding of probability and non-probability sampling,

(ii) various methods of sampling techniques

(iii) how education managers can make use of probability sampling in administration, governance, selection and for decision making in education.

6.0 Tutor Marked Assignment

i. Differentiate between systematic sampling and sequential sampling

ii. Write short notes on simple random sampling and stratified random sampling

iii. Under what circumstances should we use stratified random sampling

iv. Discuss the importance of sample size

v. Differentiate between judgment and quota sampling.

vi. A sample size of 5 is drawn without replacement from a population of size 41. If the population Standard deviation (SD) is 6.25, find the Standard Error(SE) of the sample.

7.0 References/ Further Readings

Adewoye, S. O. (2004). Basic Statistics for Engineering, Economics and Management. Lagos: Olukayode Ojo Commercial Enterprises.

Clark, G.M. and Cooke, D. (2004). Basic Course in Statistics (5th ed) New York. Oxford University Press Inc.

Levin, R.I. and Rubin, D.S. (1994). Statistics for Management (7th ed) New Jersey: Prentice Hall Inc.

Monga, G.S. (2007). Mathematics and Statistics for Economics (Second Revised Edition) New Delhi, Vikas Publishing house PVT Ltd.

Salami, K.A. (1999). Descriptive Statistics for Beginners. Oyo. Odumatt Press and Publishers.

Salami, K.A. (2001). Introduction to Basic Statistics in Management. In Adeyanju A. (ed) Introduction to Educational Management, Oyo Green Light Press and Publishers.

Salami, K.A. (2001). Basic Statistics and Data Processing in Education. In Adeyanju A. (ed) Introduction to Educational Management, Oyo Green Light Press and Publishers.

Unit 2 PROBABILITY THEORY AND DISTRIBUTION

CONTENTS

1.0 Introduction

The fact remains that chance and luck play an important role in human lives. In fact, such events as marriages, deaths, appointments, playing games etc. depend substantially on chance. For example, the birth of an individual depends on a biological chance resulting in a male or female, a white or a black, a genius or an idiot. The concept of probability is so important to every administrator and education manager to the extent that without its knowledge, human understanding of the world around him may be difficult. This unit will therefore introduce you to some relevant concepts in probability theory and distribution, the laws governing probability theory as well as conditional and unconditional probability cases.

2.0 Objectives

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

(i) understand the basic concepts associated with probability theory and distribution

(ii) identify the laws governing the operations of probability.

(iii) conceptualize the conditional and unconditional probability cases.

(iv) understand that every event or happening in this world depends on chance

(v) acquaint themselves with the formulas associated with Binomial, Poisson and Hyper- geometric distribution

(vi) solve problems with these distributions

(vii) .

3.0 Main Content

3.1 Relevant Concepts

(i) **Event:** An event is said to have occurred in an experiment if the result is a specified outcome E. If this specified outcome is observed when the experiment is performed, then, E is said to have occurred.

(ii) **Trial:** A procedure or an experiment to collect any statistical information is called a trial.

(iii) **Exhaustive Events:** A set of events is exhaustive if all possible events associated with a trial are included in the set. For example, the set of events (Head or Tail) exhausts all possibilities in the toss of a fair coin since nothing else can occur. In the roll of a die (1, 2, 3, 4, 5, 6) is an exhaustive set of events.

(iv) **Favourable Events:** Such cases as result in the happening of an event are said to be cases favourable set of events. If in tossing a fair coin, someone prior opted for Head and it actually surface, then we say it is a favourable event.

(v) **Elementary or Simple Events:** Such event that contains only a single outcome is called elementary event.

(vi) **Impossible Events:** An event which can never occur is called impossible event.

(vii) **Certain Events:** An event which is certain to occur is a certain event. For example, death to all living being is certain.

(viii) **Complementary Events:** Events E_1, E_2 are complementary if whenever E_1 occurs, E_2 does not and verse versa. For example, occurrences of head and tail in the toss of a fair coin are complementary events. The complement of an event E is written $\bar{E}$.

3.2 Fundamental Laws of Probability

(i) The Law of Addition of Probability

(a) The probabilities of the union event $E_1 + E_2$ is given by

$$P(E_1 + E_2) = P(E_1) + P(E_2) \quad P(E_1 E_2)$$

(b) If E_1, E_2 are mutually exclusive events, i.e. $P(E_1 E_2) = 0$

then, $P(E_1 E_2) = P(E_1) + P(E_2)$

(c) If E_1, E_2, E_3 are any three events

$$P(E_1 + E_2 + E_3) = P(E_1) + P(E_2) + P(E_3) \quad P(E_1 E_2) \quad P(E_2 E_3)$$

$$P(E_3 E_1) + P(E_1 E_2 E_3)$$

(d) If $E_1 E_2 E_3$ are mutually exclusive events, then,

$$P(E_1 + E_2 + E_3) = P(E_1) + P(E_2) + P(E_3)$$

(ii) The Law of Multiplication of Probabilities

a. The probabilities of the intersection events $E_1 E_2$ is given by

$$P(E_1 E_2) = p(E_1)\, p(E_2/E_1)$$

$$= P(E_2)\, p(E_1/E_2)$$

b. If $E_1 E_2$ are independent

$$p = (E_1 E_2) = p(E_1)\, p(E_2)$$

For three events E_1, E_2, E_3

$$P(E_1 E_2 E_3) = p(E_1)\, p(E_2/E_1)\, p(E_3/E_1)$$ and so on

3.3 Sample Space and Sample Point

(i) Sample Space

A set S of all possible outcomes of some given experiment is called sample space. For example, if a die is rolled once, the sample space will be all possible outcomes i.e. (1,2,3,4,5,6).

(ii) Sample Point

A particular outcome, i.e. an element in S, is called a sample point or sample. In the experiment of rolling a die once, a particular outcome, say a 4 occurring or a 2 occurring is called a sample point.

3.4 Finite Probability Spaces

Let S be a finite sample space

$S = (a_1, a_2 \quad , a_n)$

A finite probability space is obtained by assigning to each point $a_1 \quad S$ a real number P_1, called the probability of a_1, satisfying the following properties.

(i) each P_1 is non-negative, $P_1 > 0$

(ii) the sum of the P_1 is one, that is, $P_1 + P_2 + \quad P_n = 1$

Example 2.1

Three horses A, B, and C are in a race; A is twice as likely to win as B and B is twice as likely to win as C. What are their respective probabilities of winning i.e.

P (A), P (B) and P (C)?

Solution

(i) Let P (C) = p

Since B is twice as likely to win as C

P (B) = 2p

and since A is twice as likely to win as B

P (A) = 2p (B)

= 2 (2p) = 4p

Note that the sum of the probabilities must be 1. Hence,

$$P = 2p + 4p = 1 \text{ or } 7p = 1 \text{ or } P = \frac{1}{7}$$

Accordingly, $P(A) = 4p = \frac{4}{7}$

$P(B) = 2p = \frac{2}{7}$

and $P(C) = p = \frac{1}{7}$

3.5 Equiprobable Spaces

(i) A finite probability space S, where each sample point has the same probability is called an equiprobable space. If A is an event, then the probability of A denoted by P (A) would be.

$$P(A) = \frac{\text{number of element in A}}{\text{number of element in S}}$$

or

$$P(A) \quad \frac{\text{number of ways that the event A can occur}}{\text{number of ways that the sample space S can occur}}$$

Example 2.2

Let a card be selected at random from an ordinary deck of 52 cards. Let A = (the card is a space) and B = (the card is a face card).

We shall compute P (A), P (B) and P (A $\cap$ B)

Since we have an equiprobale space

$$P(A) = \frac{\text{number of spaces}}{\text{number of cards}} \quad \frac{13}{52} = \frac{1}{4}$$

$$P(B) = \frac{\text{number of face cards}}{\text{number of cards}} = \frac{12}{52} = \frac{3}{13}$$

3.6 Theorems on Finite Probability Spaces

(i) Theorem 1

The probability function P defined on the class of all events in a finite probability space satisfies the following assumptions.

Property 1 for every event B, $0 \leq P(B) \leq 1$

Property 2 P (S) = 1

Property 3 If event A and B are mutually exclusive

Then, P (A $\cup$ B) = P (A) + P (B)

(ii) Theorem 2

If $\emptyset$ is the empty set and A and B are arbitrary events, then:

P ($\emptyset$) = 0

P (A^c) = 1 P (A)

P (A/B) = P (A) P (A $\cap$ B)

i.e. P (A $\cap$ B) = P (A) P (A $\cap$ B)

Theorem 3

For any event A and B

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Corollary

For any events A, B, and C

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

Example 2.3

In a class containing 100 students, 30 attempted question A, 20 attempted managerial economic question and 10 attempted question A and managerial economics question. If a student is selected at random, find the probability P that he is taking question A or managerial Economics question

Solution

Let A = (Student taking Q/A)

and B = (student taking managerial economics)

Then $A \cap B$ = (student taking Q/A and M.E)

Since the space is equiprobable, then

$$P(A) = \frac{30}{100} = \frac{3}{10}$$

$$P(B) = \frac{20}{100} = \frac{1}{5}$$

$$P(A \cap B) = \frac{10}{100} = \frac{1}{10}$$

Then $P = P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$= \frac{3}{10} + \frac{1}{5} - \frac{1}{10} = \frac{3}{5}$$

3.7 Unconditional Probability

Let us assume there are events A, B and C whose probabilities are o.6, 0.2 and 0.4 respectively. If we are informed that event A or B happened but no categorical statement about which event happened. What is the probability that it was event A that happened?

Solution

First, we shall calculate the compound event A or B happening.

$P = P(A \cup B) = P(A) + P(B)$

$= 0.6 + 0.2$

$= 0.8$

Secondly, we need to calculate a conditional probability for A the proportion of

$P = P(A \cup B)$ supplied by P (A) i.e. $P = P(A \cup B) = 0.8$

However, 0.6 of this total came from P (A), so there is 0.6/0.8 probability that A occurred.

The value of P (A), given that B has happened, is 0.6/0.8 = 0.75.

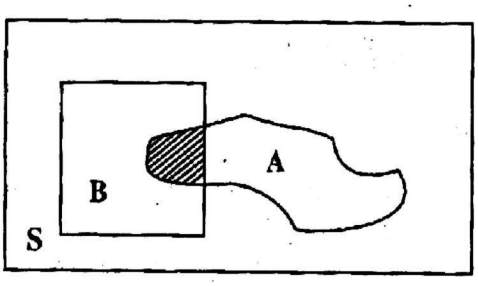
3.7.1 Conditional Probability

Conditional probabilities are often written in a shorthand format as P (x/y). this is usually read as the probability of x given that y has happened.

3.8 Conditional Probability and Independence

We shall demonstrate this with an illustration. Let B be an arbitrary event in a sample space S with P (B) > 0. The probability that an event A occurs once B has occurred or, in other word, the conditional probability of A given B, is written P (A/B), and defined as follows:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$



Theorem I

If S is an equiprobable space and A and B are events of S, then,

$$P(A/B) = \frac{\text{number of elements in } (A \cap B)}{\text{Number of element in B}}$$

or

$$P(A/B) = \frac{\text{number of ways A and B can occur}}{\text{number of ways B can occur}}$$

Example 2.4

Let a pair of dice be tossed. If the sum is 6, find the probability that one of the dice is a 2, that is,

B = (sum is 6)

= (1,5), (2,4), (3,3), (4,2), (5,1)

A = (a 2 appears on at least one die)

then, we want to find P (A/B)

Now B consist of fire elements and two of them, (2,4) and (4,2) belong to A

Therefore, $A \cap B = \{(2,4), (4,2)\}$

Then $P(A/B) = \frac{2}{5}$

3.9 Multiplication Theorem for Conditional Probability

(i) Theorem

$$P(A/B) \times \frac{P(A \cap B)}{P(B)}$$

$$P(A \cap B) = P(B)\, P(A/B)$$

We can extend this theorem by induction as follows

(ii) Corollary

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)\, P(A_2/A_1)\, P(A_3/A_1 \cap A_2) \ldots P(A_n/A_1 \cap A_2 \cap \ldots \cap A_{n-1})$$

Example 2.5

A lot contains 12 items of which 4 are defective. Three items are drawn at random from the lot one after the other. Find the probability (P) that all the three are non-defective.

The probability that the first item is non-defective is $\frac{12}{9}$ (that is, 12 items are good and 4 items are bad 12 4 = 8).

Since 8 of 12 items are non-defective, if the first item is non-defective, then defective is $\frac{7}{11}$ since 7 of the remaining 11 items are non-defective. If the first two

items are non-defective, then the probability that the last item is non-defective is $\frac{6}{10}$ since only 6 of the remaining 10 items are now non-defective. This by multiplication theorem.

$$P = \frac{8}{12} \cdot \frac{7}{11} = \frac{6}{10} \quad \frac{14}{55}$$

Self Assessment Exercise

If a bag contains 24 items of which 8 are defective, three items are drawn at random from the bag one after the other. Find the probability (P) that all the three are non-defective.

3.10 Independence

An event B is said to be independent of an event A if the probability that B occurs is not influenced by whether A has or has not occurred. In other words, if the probability of B equals the conditional probability of B given A: P (B) = P (B/A).

Formal Definition

Events A and B are independent if

$$P(A \cap B) = P(A)\, P(B)$$

Otherwise the two events are dependent.

3.11 Some Properties of Binomial Distribution

(a) Mean = $\bar{x}$ = np

(b) Variance^2 σ^2 = $\sqrt{npq}$

Example 2.6

The probability that a part one student in X. Z University will graduate is 0.6. Let us determine that out of 6 students.

(a) none will graduate

(b) one will graduate

(c) At least, one will graduate

(d) exactly two will graduate

(e) not more than two will graduate

(f) at most two will graduate

Solution

n = 6, p = 0.6, q = 1 p or 1 0.6 = 0.4

(a) P (none will graduate)

P (x = 0) $= {}^{6}C_{0}\ (0.6)^{0}\ (0.4)^{6\ \ 0}$

= 0.004096 or 0.004

(b) P (one will graduate) = P (x = 1)

P (x=1) $= {}^{6}C_{1}\ (0.6)^{1}\ (0.4)^{6\ \ 1}$

$= {}^{6}C_{1}\ (0.6)^{1}\ (0.4)^{5}$

= 6 x 0.6 x 0.01024

= 0.03686 or about 0.037

(c) P (at least one will graduate) = 1 (none will graduate) or p (x>1) = p (x=1) + p (x=2) + p (x=3) + p (x=4) + p (x=5) p (x=6)

Alternatively

P (x≥1) = 1 p (x<1)

1 p (x = 0) = 1 0.004)

= 0.996

P (x≥1) = 0.996

d. P (exactly two will graduate) = P (x = 2)

$= {}^{6}C_{2}\ (0.6)^{2}\ (0.4)^{6\ \ 2}$

$= {}^{6}C_{2}(0.6)^{2}\ (0.4)^{4}$

= 15 x 0.36 x 0.0256

P (x=2) = 0.13824

e. P (not more than two will graduate) = P (x ≤ 2)

P (x ≤ 2) = P (x = 0) + P (x = 1) + P (x = 2)

$= {}^{6}C_{0}\ (0.6)^{0}\ (0.4)^{6} + {}^{6}C_{1}\ (0.6)^{1}\ (0.4)^{5} + {}^{6}C_{2}\ (0.6)^{2}\ (0.4)^{4}$

= 0.00409 + 0.0368 + 0.13824

P (x ≤ 2) = 0.17913

f. P (at most two will graduate) = p (x ≤ 2)

P (x ≤ 2) = P (x = 0) + P (x = 1) + P (x = 2)

Ans = 0.17913

Example 2.7

A manufacturer of spare parts that are needed in electronic device guarantee that a box of its part contains at most two defectives parts. If the box holds 20 parts and experience has shown that the manufacturing process produces 2 percent defective items. What is the probability that a box of the part will satisfy the guarantee.

Solution

X = 0, 1, 2, N = 20 $P = \frac{2}{100} = 0.02$

Guarantee = $P (x \leq 2)$

$4\ P (x \leq 2) = P (x = 0) + P (x = 1) + P (x = 2)$

$= {}^{20}C_0 (0.02)^0 (0.98)^{20} + {}^{20}C_1 (0.02)^1 (0.98)^{19} + {}^{20}C_2 (0.02)^2 (0.98)^{18}$

$= 0.668 + 0.272 + 0.053 = 0.993$

$P (x \leq 2) = 0.993$

Self Assessment Exercise 1

1. What is the probability of getting at least 4 head in 6 tosses of a fair coin.

3.2 Poisson Distribution Formula

$P (x) = \frac{\lambda \ e \ \lambda}{X!}$

where X = 0_1, 1, 2, 3 ... n

e = 2.71828

$\lambda = np$

3.2.1 Properties of the Poisson Distribution

(a) Mean = $\mu = \lambda$

(b) Variance $\sigma^2 = \lambda$

(c) Standard deviation = $\sigma = \sqrt{\lambda}$

Example 2.8

If the probability that an individual suffers a bad reaction from injection of a given serum is 0.003, Let us determine that out of 2000 individuals.

(a) Exactly 3 individuals will suffers a bad reaction

(b) More than 2 individuals will suffer a bad reaction

Solution

P = 0.003, n = 2000, λ = np = 6

(a) $P(x = 3) = \frac{6^3 e^{-6}}{3!}$

$= \frac{216}{3!e^6}$

or $\frac{216}{6x(2.7182)^6} = 0.08925$

P (x = 3) = 0.08925

Alternatively, you may calculate $\frac{6^3 e^{-6}}{3!}$ directly with your calculator and get 0.08923

(b) P (x > 2) = P (x = 3) + P (x = 4) + + P (x = 2000)

P (x > 2) = 1 P (x = 0) + P (x = 1) + P (x = 2)

$= 1 - \left[\frac{6^0 e^{-6}}{0!} + \frac{6^1 e^{-6}}{1!} + \frac{6^2 e^{-6}}{2!} \right]$

= 1 (0.002378 + 0.01487 + 0.04462)

= 1 (0.61968)

P (x > 2) = 0.93803 or 93.8 percent

Example 2.9

Suppose a customer arrives at a service station according to a poisson distribution and that the arrival average is 24 per hour. What is the probability of number arrival in a five minute interval?

Solution

$P = \frac{24}{60}$, $N = 5$, $x = 0$, $\lambda = Np = 5 \times \frac{24}{60}$

$\lambda = 2$, Hence $P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$

$$P(x) = \frac{2^{0}}{0!\,(2.71828)^{2}}$$

$P(x) = 0.135$

Self Assessment Exercise 2

1. If 3% of the electric bulb manufactured by a company is defective, find the probability that in a sample of 100 bulb, (a) 0, (b) 2 bulbs will be defective.

3.3 Hypergeometric Distribution Formula

$$f(y) = \frac{\binom{n_1}{y_1}\binom{N-n_1}{n-y}}{\binom{N}{n}}$$

Where $y = 0, 1, 2, \ldots n$

Example 2.10

A young boy went fishing and caught 10 fishes with blue gills of which 4 were under 6 inches in length. When he returned home, the pulled 2 fishes from the pail at random and without replacement so that each fish has the same probability of been drawn.

If y denote the number of fish under 6 inches in the sample. Define the p.d.f. of random variable of y.

Solution

$n_1 = 4, N = 10, n = 2$

$$f(y) = \frac{\binom{4}{y}\binom{6}{2-y}}{\binom{10}{2}} \qquad y = 0, 1, 2$$

When y = 0

$$f(0) = \frac{\binom{4}{0}\binom{6}{2}}{\binom{10}{2}}$$

$$= \frac{1 \times 15}{45}$$

$$= 0.333$$

When y = 1

$$f(1) = \frac{\binom{4}{1}\binom{6}{1}}{\binom{10}{2}}$$

$$= \frac{4 \times 6}{45}$$

$$= 0.533$$

Where y = 2

$$f(2) = \frac{\binom{4}{2}\binom{6}{2}}{\binom{10}{2}}$$

$$= \frac{6 \times 1}{45}$$

$$= 0.133$$

Example 2.11

A shipment of 20 tape recorders contains 5 that are defective. If 10 of them are randomly chosen for inspection, what is the probability that 2 of the 10 will be defective?

Solution

$N = 20, \quad n = 10 \quad n_1 = 5 \quad y = 2$

$$f(y) = f(2) = \frac{\binom{5}{2}\binom{15}{8}}{\binom{20}{10}}$$

$$= 0.348$$

Self Assessment Exercise

A shipment of 100 tape recorders contains 25 that are defective. If 10 of them are randomly chosen for inspection, what is the probability that 2 of the 10 will be defective?

4.0 Conclusion

In this unit, students have been exposed to relevant concepts in probability theory and distribution. Students have also learnt the fundamental laws of probability. Again, several theorems relating to many aspects of probability were discussed and explained with examples. Conditional and unconditional probability cases were equally treated. Furthermore, students were adequately enlightened through this unit that all human trials are based on probability or chance.

5.0 Summary

This unit has enlightened students on:

i. relevant concepts in probability theory and distribution
ii. fundamental laws guiding probability theory
iii. several theorems relating to probability
iv. conditional and unconditional probability cases
v. finite probability spaces.

6.0 Tutor Marked Assignment

i. differentiate between certain and impossible events

ii. define the concept sample space

iii. A salesman for a company sells two products, A and B. During the morning he makes three calls to customers. Suppose the chance that on any call he makes a sale of product A is 1 in 3 and the chance that he makes a sale of product B is 1 in 4. Suppose also that the sale of product A on any call is independent of the sale of product B and that the results of the three calls are independent of one another. What is the probability that the salesman will:

(a) sell both products, A and B at the first call.

(b) sell one product at the first call

(c) make no sales of product A during the morning

(d) make at least one sale of product B during the morning

(e). The number of plant of a certain species at a particular site is known to have a poisson with mean of 2 plant per metre squared. Find the probability of

(i) exactly 2, (ii) more than 2 (iii) less than 2 plants per metre squared

(f) Find the probability that in a family of 4 children, there will be (a) at least 1 boy (b) at least 1 boy and 1 girl.

7.0 References/ Further Readings

Adewoye, S. O. (2004): Basic Statistics for Engineering, Economics and Management, Lagos: Olukayode Ojo Commercial Enterprises.

Clarke, G. M. and Cooke, D. (2004): A Basic Course in Statistics (5th Ed.) New York. Oxford University Press.

Hildebrand, D.K. & Lyman, O. (1998). Statistical thinking for managers (4th Ed.) California: Duxbury Press.

Levin, R.I. &Rubin, D.S. (1997). Statistics for Management (7th Ed.) New Jersey: Prentice Hall International.

Monga, G.S. (2007). Mathematics and Statistics for Economics (Second Revised Edition) New Delhi. VIKAS Publishing House PVT Ltd.

Salami, K.A. (1999). Descriptive Statistics for Beginners. Oyo Odumatt Press and Publishers.

Salami, K.A. (2001). Introduction to Basic Statistics in Management. In Adeyanju, A. (Ed.) Introduction to Educational Management. Oyo Green Light Press and Publishers.

Unit 3 ESTIMATION

CONTENTS

1.0 Introduction

Managers and administrators of educational system are daily engaged in making decision for their organizations and such decisions are usually marked by uncertainty because the available information, in most cases, is either incomplete or inadequate. In such situation, managers have to resort to some kind of estimation which may be guesswork, or something based on experience or rather, something evolved by the use of scientific method. This unit will therefore teach you how to use statistical techniques to infer or estimate the parameter of a population with the help of the characteristics of a sample selected from the population. Estimation of a good sample randomly selected from a population is primarily the focus of this unit.

2.0 Objectives

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

i. differentiate between estimation and estimator

ii. describe point and interval estimation

iii. explain how estimation is used for decision making

iv. list the properties of a good estimator

v. identify various types of unbiased estimators.

3.0 Main Content

3.1 Estimation Defined

Estimation is a rough figure or likely total figure of observations, events or objects. This is usually carried out with the help of information obtained from a sample which is drawn from a population. For instance, if we calculate a statistic, say the arithmetic mean from a sample. This statistic may be inferred to as approximation of the corresponding population parameter

3.1.1 Estimator Defined

An estimator is a statistic used to estimate a population parameter. It is a function of the sample observations.

3.2 Types of estimates

3.2.1 Definition of Point Estimate

Point estimate is an estimate of a population parameter given by a single number. In other words, point estimate is a single numerical value used as an estimate of a population parameter**.** The sample statistics, such as , $s_{\bar{y}}$ and , $\bar{p}$ that provides the point estimate of the population parameter is called point estimator.

3.2.2 The properties of a good point estimator are:

i. unbiasness

ii. consistency

iii. efficiency

i. unbiasness: If it is an unbiased estimator of the parameter , θ then, E (t) = . A $\emptyset$ sample statistic $\bar{x}$ is an unbiased estimator of the corresponding population parameter μ. The expected value of the sampling distribution of the statistic equals the corresponding population parameter. Thus, if the statistic t, which is a function of sample value is an unbiased estimator of the population parameter

μ then E (t) = μ if E (t) μ. Then, the estimatort is biased. We can also measure it by writing Bias = E (t) - μ.

ii. **Consistency:** If a statistic t comes closer and closer to the population parameter as the sample size becomes large, that is t → as n then, t is said to be a consistent estimator of . Consistency is a limiting property based on the sample size n. An estimator usually becomes consistent as n increases. This implies that an increase in the sample size n goes on adding information about the population parameter. Symbolically, we can write n , x μ, meaning that the sample mean becomes a consistent estimator of the population mean μ.

iii. **Efficiency:** There can be several consistent estimators, say t, t_2, t_3 for the same population parameter . The most efficient or the best estimator of these is the one which has the least variance (i.e. least standard error).

If v (t_1) < v(t_2) < (t_3) or if SE (t_1) < SE (t_2) < SE (t_3) then t_1 is more efficient than t_2 and t_3 is more efficient than t_3. Relating efficient estimator to the school system is like when a school manager wants to find out which method will be more efficient in teaching a particular topic in a subject. If after treatment, he found out that participatory method < team method < lecture method. The manager will then accept participatory method as the most efficient method, although other methods are also good and result oriented.

3.2.3 Definition of Interval Estimate

An interval estimate of a population parameter provides an interval believed to contain the value of the parameter. Interval estimate places the unknown parameter between two limits. Let us consider the following examples.

i. 61kg 75kg for the mean weight μ of adult females.

ii. 0.50 0.60 for the proportion of eligible voters supporting the chairman of L. G.

iii. 58 60 litre for the mean petrol μ

iv. N52,200.00 64,400.00 for the mean wage for graduate teacher

3.4 Estimation in Decision Making

As earlier said in the introduction of this unit, managers and administrators of schools do engage themselves daily in making decision for their organization under a condition of uncertainty as a result of incomplete and inadequate information. Since there should be no vacuum in management and decision making, managers and administrators often resorts to some kind of estimation usually based on either guesswork, managers long experience at work or sometimes based on scientific methods.

Whatever method a manager adopts, the basic fact remains that a decision must be made and any estimate arising from such decision is expected to influence the working of large or small organizations.

3.5 Mean Square Error.

The performance of an estimator is measured by

$$MSE = E(t - \text{ø})^2$$

$$= Var.(t) + (E(t) - \text{ø})$$

$$= Var.(t) + B^2$$

where the bias $B = E(t) - \text{ø}$

In the case of the sample mean used as an estimator of the population mean;

$B = E(\bar{x}) - \mu = \mu - \mu = 0$

$MSE = E(\bar{x} - \mu)2 = var.(x) = \frac{2}{N}$

Self Assessment Exercise 1

i. Give accurate definition of interval estimate.

ii. Give example of point estimate.

3.6 Confidence

The degree of confidence in term of percentage attached to an interval estimate implies the extent of the confidence that the population parameter is included in the range of the interval estimate. This is expressed in terms of a probability statement to indicate the degree of confidence in the estimate interval.

3.6.1 Confidence Interval Estimate and probability Statement

The area under the normal curve between -1.96 SE and +1.96 SE of the population mean is known to be 0.95. We can describe this by saying that if the sample data are normally distributed, the probability of the mean (average) lying between ± 1.96 SE is 0.95. In other words, the chance of the population mean lying between ± 1.96 is 95% or 0.95. Again, we can say that our confidence is 95% that the interval contains the population mean. You will re-call that in the previous units before this, we asked you to assume the existence of a population and that because we cannot reach all the population or calculate the mean of the population, we resort to taking a representative sample randomly selected from the assumed population. And that as the sample selected increases, its mean tends to the population mean.

Therefore, the probability (in the above case 0.95 or 95%) associated with an interval estimate is the level of confidence. Below are the types of confidence interval usually employed in statistics;

i. 95% corresponding to ± 1.96 SE

ii. 95.45% corresponding to ± 2 SE

iii. 99% corresponding to ± 2.58 SE

iv. 99.73% corresponding to ± 3 SE

Other confidence level can be employed as required by using the normal table. For instance the 90% confidence interval correspond to ± 1.94 SE.. The value corresponding to ± 1.94 gives the upper confidence limit (U C L) and that which corresponds to -1.64 SE gives the lower confidence limit (L C L).

Note that the wider the confidence limit, the greater is the certainty of the statement. However we need to be careful so that the interval chosen will not be too wide for statistical operation.

Exercise 1. From a large consignment of chalk supplied to the school, a sample of 100 packets was selected and found to contain 10% bad ones. Fine 95.45% and 99.73% confidence limit for the bad ones.

$P = 0.1,\ q = 0.9\ SE(p) = \sqrt{pq} = 0.1\sqrt{\frac{x\ 0.09}{n}} = \frac{0.015}{400}$

where P, Q are not known, so we used r, q.. The 95.45% limits are 0.1 ± 2x 0.015 = 0.13, 0.07 the 99.73% limits are 0.1 ± 3x 0.015 = 0.145, 0.055. We shall discuss more of this in the next unit when testing hypothesis

4.0 Conclusion

In this unit, we have seen how education managers take or make decisions under uncertainty as a result of inadequate information or data base. The role of estimate derived from good sample randomly selected from a given population in deciding population parameter was discussed The properties of a good estimate were adequately highlighted on this unit. How to calculate mean square error was also explained.

5.0 Summary

In this unit, you have identified;

i the difference between the concepts of estimation and estimator

ii. the role of estimation in decision making

iii. the properties of good estimators

iv. types of unbiased estimators

v. how to calculate mean square error

6.0 Tutor Marked Assignment

Table below shows the weights of 100 male students at EACOED Model High School, Oyo. A random sample representing the weights of all the 100 students at the Model High School.

1. Weight of 100 students randomly picked

Weight (kg)	60 - 62	63 - 65	66 - 68	69 - 71	72 - 74	**Total**
No of Student	5	18	42	27	8	100

Action Required

Determine the unbiased and efficient estimates of

(a) the true mean

(b) the true variance

2. Discuss two properties of a good point estimate.

7.0 References/ Further Readings

Adewoye, S. O. (2004). Basic Statistics for Engineering, Economics and Management. Lagos: Olukayode Ojo Commercial Enterprises.

Hildebrand, D.K. & Lyman, O. (1998). Statistical thinking for managers (4th Ed.) California: Duxbury Press.

Levin, R.I. &Rubin, D.S. (1997). Statistics for Management (7th Ed.) New Jersey: Prentice Hall International.

Monga, G.S. (2007). Mathematics and Statistics for Economics (Second Revised Edition) New Delhi. VIKAS Publishing House PVT Ltd.

Salami, K.A. (1999). Descriptive Statistics for Beginners. Oyo Odumatt Press and Publishers.

Salami, K.A. (2001). Introduction to Basic Statistics in Management. In Adeyanju, A. (Ed.) Introduction to Educational Management. Oyo Green Light Press and Publishers.

Unit 4 TESTING OF HYPOTHESIS

CONTENTS

1.0 Introduction

This unit will teach you how to use all the statistical tools we have been dealing with right from unit one. To do this, this unit therefore concentrates on how to arrive at a decision through the use of hypothesis testing. This involves the rejection or acceptance of a null hypothesis.

2.0 Objectives

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

i define what a hypothesis is

ii construct null and alternative hypotheses

iii identify type I and II errors in decision making

iv identify critical or acceptance region

v describe penalty for committing either type I or II error

vi understand decision rule in hypothesis testing

vii. explain the use of student t-statistic

3.0 Main Content

3.1 Basic Definitions on Tests of Hypothesis

3.1.1 Statistical Decisions

When we make a decision about population on the basis of sample information, then, we have statistical decisions e.g. we may decide on the basis of sample data whether a particular teaching method is effective or not.

3.1.2 Statistical Hypothesis

When we make assumptions or guesses about the population (such assumptions may or may not be true), then, we have statistical hypothesis. Usually, a statistician would formulate a statistical hypothesis for the sole purpose of rejecting or nullifying it.

3.2 Hypothesis Testing

Testing a statistical hypothesis on the basis of a sample will enable us to decide whether the hypothesis should be accepted or rejected. The procedure which, on the basis of sample result, enables us to decide whether a hypothesis is to be accepted or rejected is called Hypothesis testing or Test of Significance.

3.3 Acceptance of a hypothesis

The acceptance of a hypothesis implies that there is no evidence from the sample that we should believe otherwise.

3.3.1 Rejection of a Hypothesis

The rejection of a hypothesis leads us to conclude that the hypothesis is false. This way of putting problem is convenient because of the uncertainty inherent in the

problem. In view of this, we must always briefly state the hypothesis that we hope to reject

A hypothesis stated in the hope of being rejected is called a null hypothesis and is denoted by H_o. For example, if we are interested in comparing the performances of a group of students in Economics and Mathematics, we may formulate the null hypothesis as follows;

> ***Ho:*** *There is no significant difference in the mean scores of students in Economics and Mathematics.*

However, if H_o is rejected, it may lead to the acceptance of an alternative hypothesis denoted by H_1. H_1 may be stated as follows;

> *There is significant difference in the mean scores of students in Economics and Mathematics.*

3.4 Type I and II Errors in statistics

The following are the ways by which we can commit errors when accepting or rejecting hypothesis.

i. Type I Error is when we reject a true hypothesis.

ii. Type II Error is when we accept a false hypothesis.

iii. The other true situations are desirable, that is when we accept a true hypothesis and when we reject a false hypothesis.

In a tabular form

	Accept H_o	Reject H_o
H_o	Accept true H_o	Reject true H_o
True	Desirable	Type I Error
H_o	Accept false H_o	Reject false H_o
False	Type II Error	Desirable

Self Assessment Exercise 1

a When can a manager commit Type one and Type two Error

b Differentiate between null hypothesis and Alternative hypothesis

3.5 Procedure for Carrying out Tests of Hypothesis

Below are the general steps involved in testing Hypothesis.

Step 1: State the null hypothesis (H_0) and the alternative hypothesis (H_1)

Step 2: State the criterion level of significance (if they are not given).

Step 2b: State your decision rule.

Step 3: Calculate the mean and standard deviation of the given population or their estimates (if they are not given)

Step 4: Compute the appropriate statistic, for example z or t value using the appropriate formula and get the calculated value.

Step 5: Determine the tabulated or critical value corresponding to the given level of significance (there is a table already prepared for this)

Step 6: Make your conclusion after comparing the calculated value with the tabulated value based on stated decision rule in step 2b

3.5.1 Null and Alternative Hypothesis

The hypothesis that is formulated is called the Null hypothesis and is denoted by H_o. This is to be tested against other possible stated alternative hypothesis. The alternative hypothesis is usually denoted by H_1.

The null hypothesis implies that there is no significant difference between the statistic and the population parameter. To test whether there is no significant difference between the sample mean ( ) and the population mean (μ) we write the null hypothesis as; H_o: $\bar{x} = \mu$.

The alternative hypothesis would be H_1: $\bar{x} \neq \mu$. This means $> \mu$ or $< \mu$, this is called a two tailed hypothesis.

The null hypothesis could also be between two sample means such as $\bar{x}_1 = \bar{x}_2$ or $\bar{x}_1 - \bar{x}_2$. If we are interested in comparing the means of two independent scores say male and female scores in mathematics, the null hypothesis would read thus:

> ***H_o:*** *There is no significant difference between the means scores of male and female students in mathematics.*

While the alternative hypothesis would also read

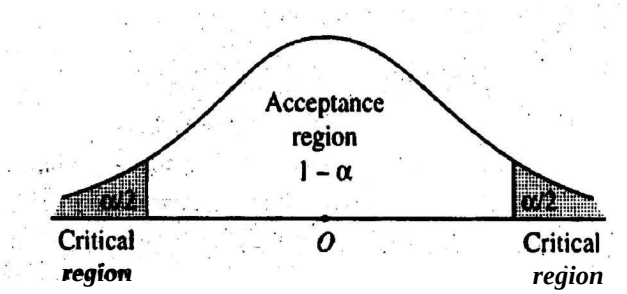
> ***H_1:*** *There is significant difference between the mean scores of male and female students in mathematics.*

3.6 Critical Region (Region of Hypothesis Rejection)

Let us consider test problems arising out of Type I error. The level of significance of test is the maximum probability with which we are willing to take a risk of Type I error. If we take a 5% significance level (p = 0.05) we are 95% (p = 0.95) that a right decision has been made.

A 1% significance level (p = 0.01) makes us 99% confident (p = 0.99) about the correctness of the decision.

The critical region is the area of the sampling distribution on which the test statistic must fall for the null hypothesis to be rejected. We can say that the critical region corresponds to the range of values of the statistic which according to the test requires the hypothesis to be rejected.



Two Tailed Test

If the level of significance is α, then in a two tailed test, a part of the critical region (say half, i.e. $\alpha/2$) is placed on the right and another part (i.e. $\alpha/2$ on the left as shown in the graph above

3.7 Penalty

Let us examine the type of error that is bad. Usually type II error is considered the worse of the two, though, it is usually the circumstances of a case that decid e. You will recall that type I error is accepting the hypothesis that a guilty person is innocent and type II error is accepting the hypothesis that an innocent person is guilty then, type II error would be dangerous to commit. The penalties and costs associated with an error determine the balance or trade off between type I and type II errors.

Usually type I error is shown as the shaded area; say 5% of a normal curve which is suppose to represent the data. If the sample statistic, say the sample mean, falls in the shaded area, the hypothesis is rejected at 5 per cent level of significance.

3.8 Standard Error

The concept of standard error in statistic is used to test the precision of a sample and provides the confidence limits for the corresponding population parameter.

The statistic may be the sample arithmetic mean , the sample proportion p etc.

The standard error (SE) of any such statistic is the standard deviation of the sampling distribution of the statistic. Given below is SEs commonly used in statistics. The number of observation is n.

Notation:

$\bar{x}$	sample mean
μ	population mean
s	sample standard deviation
	population standard deviation
p	sample proportion, q = 1 p
P	population proportion, Q = 1 p

$$SE(x) = \frac{}{n}, \qquad SE(p) = \sqrt{\frac{PQ}{n}}$$

SE of difference between two means, $\bar{x}_1$, $\bar{x}_2$ or two proportions p_1, p_2 with sample sizes n_1, n_2 will be written thus:

$$SE(x_1 \quad x_2) = \sqrt{\frac{Q_1 + Q_2}{n_1 \qquad n_1}}$$

SE (p_1 p_2) = $_1Q_1\sqrt{P}_2Q_2$ P

n_1 n

3.9 Decision Rule

Suppose μ is distributed normally with mean 0 and S.D. 1. Symbolically we write μ - N (0, 1). If the expected value of μ is written E (μ) the standardized normal variate is;

$$z = \frac{\mu - E(\mu)}{SE(\mu)}$$

3.10 How to Test for the Significance of Correlation Coefficient (r)

Let (x_1, y_2) (x_2, y_2)... (x_n, y_n) be a random sample from a normal population. To test the hypothesis that the population coefficient of correlation is zero, we use;

$$t = r\sqrt{\frac{n-2}{1}} \quad \text{with n 2 degree of freedom}$$

r^2

If /t/ > table value of t_{n-2}, 0.05 then r, is significant. If /t/ < t_{n-2}, 0.05 then r is not significant. In unit one of module 3 we got our correlation coefficient to be 0.087. The question to ask is whether r is significant for the existence of correlation in the population.

$$/t/ = 0.087\sqrt{\frac{10-2}{1-0.087^2}} = \frac{2.83}{0.992} = 0.25$$

Table value of t_8, 0.05 = 2.30.

The observed value is = 0.25

Therefore, the observed value is less than table value (0.25 < 2.30)

The coefficient of correlation is insignificant or not significant

4.0 Conclusion

This unit has taught you how to formulate and test hypothesis at 5% level of significance. Required degrees of freedom as well as the critical regions of acceptance and rejection of hypotheses were equally treated. The two common errors usually committed in statistic (Type I and II Errors) were highlighted, pointing out the avoidale type II error. Examples of null and alternative hypotheses were illustrated while the importance of standard error was explained. The unit concluded with a hypothetical example of how to calculate student t statistic.

5.0 Summary

In this unit, you have been exposed to:

i. the formulation of hypotheses for decision making
ii. the critical region of acceptance and rejection of the result of the hypothesis
iii. the two commonly committed errors in statistical analysis and the one to be avoided among the two.
iv. how to calculate student t-statistic

6.0 Tutor Marked Assignment

i. Discuss how a statistician would commit type I and II errors.
ii. Define the concept Hypothesis testing
iii. Differentiate between a null and alternative hypotheses
iv. Given two independent means and standard deviation for 2 sample sizes of 80 and 47 as

$\bar{X}_1 = 10.25$, $\bar{X}_2 = 6.44$, $SD_1^2 = 7.12$, $SD_2^2 = 4.11$

calculate the t-statistic, given the critical t as 1.96 at $p<0.05$ level of significance with df = 125. Indicate whether we should accept or reject the formulated null hypothesis which says there is no significant difference in the two means.

7.0 References/ Further Readings.

Adewoye, S. O. (2004): Basic Statistics for Engineering, Economics and Management, Lagos: Olukayode Ojo Commercial Enterprises.

Hildebrand, D.K. & Lyman, O. (1998). Statistical thinking for managers (4th Ed.) California: Duxbury Press.

Levin, R.I. &Rubin, D.S. (1997). Statistics for Management (7th Ed.) New Jersey: Prentice Hall International.

Monga, G.S. (2007). Mathematics and Statistics for Economics (Second Revised Edition) New Delhi. VIKAS Publishing House PVT Ltd.

Salami, K.A. (1999). Descriptive Statistics for Beginners. Oyo Odumatt Press and Publishers.

Salami, K.A. (2001). Introduction to Basic Statistics in Management. In Adeyanju, A. (Ed.) Introduction to Educational Management. Oyo Green Light Press and Publishers.

MODULE THREE

DATA PROCESSING IN EDUCATIONAL MANAGEMENT

\

Unit 1 THE CORRELATION ANALYSIS

CONTENTS

1.0 Introduction

In the previous units, most o f the analysis and methods we have developed and used have been for dealing with one variable only. But often than not, several characteristics are measured on each member of a sample, and it may be of great interest to ask whether the variables are interrelated. In this unit, we shall learn the methods which investigate whether two quantitative variables are related or not.

2.0 Objectives

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

(viii) compute correlation coefficient using Pearson Product Moment Correlation formula.

(ix) test the value of (r) for significance.

(x) compute rank correlation coefficient using Spearman Rank Correlation formula.

(xi) identify the coefficient of determination and that of alienation.

(xii) compute standard error of r and probable error of r.

(xiii) describe the fallacies involved in the interpretation of calculated correlation coefficients.

3.0 Main Content

3.1 The Concept of Correlation

Correlation is simply defined as the relationship between two sets of observations. It is also a statistical device to measure the amount of similarity and variations, that is, the degree of association between series of pairs of sets of observations. Correlation coefficient is a useful tool for finding how much a variable changes, corresponding to the average amount of change in other variables in the data.

3.2 Pearson s Coefficient of Correlation

The coefficient of correlation is a measure of the degree of relationship between two variables X and Y. It gives, in numerical terms, the extent to which sample observations on X are correlated with sample observations on Y.

The coefficient of correlation, denoted by r, is defined in such a way as to be pure number so that the units of X and Y do not affect the result. The value of r lies between -1 and +1 i.e. $1< r < 1$

r = +1 shows a perfect positive correlation.

r = -1 shows a perfect negative correlation.

r = 0 shows absence of correlation.

r = nearly equal to +1 shows a strong or significant positive correlation.

r = nearly equal to zero shows an insignificant correlation.

3.2.1 Correlation Formulas

There are two frequently used formulas for finding the correlation coefficient. These are Pearson Product Moment Correlation formula and Spearman Rank Order Correlation formula. Others not frequently used are: point biserial; the biserial; phi; tetrachoric; Kendall s tau and Kendall s coefficient of concordance.

3.2.2 The Pearson Formula

This is usually represented with a symbol r and can only be applied if the two sets of scores are at least at the interval level.

This coefficient can be calculated using two methods. These methods are

(i) Raw score method and (ii) Deviation method.

a The Raw Score Method.

The formula for this is written as

$$r = \frac{N\copyright XY \quad \copyright X\copyright Y}{\sqrt{[N\copyright X^2-(\copyright X)^2]\,[N\copyright Y^2 \quad (\copyright Y)^2}}$$

Example 1.1: Find the Coefficient of correlation for the following 10 sample observations.

X: 10 35 28 20 35 15 23 15 28 30

Y: 28 25 20 20 20 40 18 10 40 35

Table 1.1: Computation of Coefficient of Correlation using Raw Score Method

X	**Y**	**X^2**	**Y^2**	**XY**
10	28	100	784	280
35	25	1225	625	875
28	29	784	841	812
20	20	400	400	400
35	20	1225	400	700
15	40	225	1600	600
23	18	529	324	414
15	10	225	100	150
28	40	784	1600	1120
30	35	900	1225	1050
239	265	6397	7899	6401

Where, N = 10, $\copyright X = 239$, $\copyright Y = 265$, $\copyright X^2 = 6397$

$\copyright Y^2 = 9899$ and $\copyright XY = 6401$

By substitution into the formula, we get

$$r = \frac{10 \times 6401 \quad 239 \times 265}{\sqrt{[10 \times 6397 \quad (239)^2] [10 \times 7899 \quad (265)^2]}}$$

$$r = \frac{64010 \quad 63335}{\sqrt{[63970 \quad 57121] [78990 \quad 70225]}}$$

$$r = \frac{675}{\sqrt{(6849)\ (8765)}}$$

$$= \frac{675}{\sqrt{(82.76)\ (93.62)}}$$

$$= \frac{675}{7747.99}$$

$$r = 0.087.$$

b The Deviation Method

The formula for this is written:

$$r = \frac{\sum (x-\bar{x})\ (y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$$

Which can be re-written as

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

Where $x = (x-\bar{x})$ and $y = (y-\bar{y})$

Using the same set of scores x, y, we obtain the table below.

Table 1.2: Computation of r using Deviation method

x	Y	x-x	y-y	(x-x) (y-y)	$(x-x)^2$	$(y-y)^2$
10	28	+13.9	+1.5	-20.85	193.21	2.25
35	25	+11.1	-1.5	-16.65	123.21	2.25
28	29	+4.1	+2.5	+10.25	16.81	6.25
20	20	-3.9	-6.5	+25.35	15.21	42.25
35	20	+11.1	-6.5	-72.15	123.21	42.25
15	40	-8.9	+13.5	-120.15	79.21	182.25
23	18	-0.9	-8.5	+7.65	0.81	72.25
15	10	-8.9	-16.5	+146.85	79.21	72.25
28	40	+4.1	+13.5	+55.35	16.81	182.25
30	35	+6.1	+8.5	+51.85	37.21	72.25
239	265			+67.5	684.90	876.5

n = 10 n = 10

x = 23.9 y = 26.5

From the table, we have

©(x-x) (y-y) = +67.5

$©(x-x)^2$ = 684.90

$©(y-y)^2$ = 876.5.

Substituting the above values into the formula, we get

$$r = \frac{+67.5}{\sqrt{684.90 \times 876.5}}$$

$$r = \frac{67.5}{26.17 \times 29.61}$$

$$= \frac{67.5}{774.89}$$

r = 0.087.

The same result or coefficient with the Raw score method was obtained

3.3 Spearman Rank Order Coefficient of Correlation

This is usually represented with the Greek letter (rho) or r! and is defined by the formula

$$r^1 = 1 - \frac{6©D^2}{N(N^2-1)}$$

Example 1.2: Find the coefficient of correlation for the following 10 sample observations using Spearman Rank Order Correlation Formula.

X	10	35	28	20	35	15	23	15	28	30
Y	28	25	29	20	20	40	18	10	40	35

Table 1.3: Computation of Coefficient of Correlation using Spearman Rank Order Formula.

X	y	Rx	Ry	Rx-Ry	D^2
10	28	10	5	+5	25
35	25	1.5	6	-4.5	20.25
28	29	4.5	4	+0.5	0.25
20	20	7	7.5	-0.5	0.25
35	20	1.5	7.5	-6	36
15	40	8.5	1.5	+7	49
23	18	6	9	-3	9
15	10	8.5	10	-1.5	2.25
28	40	4.5	1.5	+3	9
30	35	3	3	0	0
					151

R = rank of each items

Substituting these values into the formula, we obtain

$$r^1 = 1 - \frac{6(151)}{10(10^2-1)}$$

$$= 1 - \frac{906}{990}$$

$$= 1 - 0.915$$

$$r^1 = 0.085$$

Self Assessment Exercise

Use both Pearson s raw score and deviation methods to find the correlation coefficient of the following sets of figures.

X	5	0	3	1	2	2	5	3	5	4
Y	1	2	1	3	3	4	3	1	0	2

3.4 Coefficient of Determination

The square of the correlation coefficient is called the coefficient of determination. The coefficient of determination which is r^2 is the proportion of total variation in y that is explained by the linear relation between x and y.

Coefficient of Determination = r^2

In formula form

$r^2 = \frac{\text{Explained variation of all items}}{\text{Total Variation of all items}}$

Example 1.3: If r = 0.6, the coefficient of determination is 0.36.

3.5 Coefficient of Alienation

The absence of relationship between two variables can be expressed by an index of lack of relationship which may be written as

Coefficient of Alienation = $\sqrt{1-r^2}$

Which is the opposite of the coefficient of correlation.

Example 1.4: If r = 0.6, the coefficient of Alienation is

$$\sqrt{1-r^2} = \sqrt{1-0.36} = \sqrt{0.64} = 0.8$$

3.6 Standard Error and Probable Error.

If r is the coefficient of correlation between a pair of values of x and y, then, the standard error designated by (SE) of r is given by

$$SE(r) = \frac{1-r^2}{\sqrt{n}}$$

While the probable error of r is given by

$$PE(r) = 0.6745 \quad SE(r) = 0.6745 \frac{1-r^2}{\sqrt{N}}$$

Example 1.5: If r = 0.75

$$SE = \frac{1-0.75^2}{\sqrt{100}} = \frac{1-0.5625}{\sqrt{100}} = \frac{0.4375}{10} = 0.04375.$$

PE (r) = 0.6745 x SE(r) = 0.04375

Ans PE (r) = 0.6745 SE (r) = 0.0295.

Note that PE (r) is the limit such that the probability of r lying between r + PE (r) is exactly 2 provided r is normally distributed, PE (r) will not be negative.

3.7 Fallacies in Interpreting Calculated Correlation Coefficient

It is good at this juncture for education managers to clearly understand that having a positive or negative correlation coefficient does not imply causes. For example, when the scores of two independent subjects like mathematics and physics are computed and found to be positively correlated. Education managers should not take for granted that good performance in mathematics caused or contributed to the good performance of the student in physics. There could be other factors other than the good performance in mathematics.

What the value of r tells us is that it measures only the strength of linear relation between x and y. Let us examine one or two examples here to show that correlation coefficient does not imply causes. The first example is that scientific researches have shown that the high blood pressure of patient is denoted by x and his heart beat/rate denoted by y. When the two was correlated, it was found to be positively correlated, but further research showed that the cause of high blood pressure was not the heart beat but the patient s weight which is represented by z.

The importance of the fallacy is that education managers should take the good performance of a teacher in the school as a result of regular payment of his salaries and allowances and other conditions of service just as ordinary linear relationship. The cause of teachers good performance may be a third variable yet unknown. Therefore, serious thought and research should be directed to know such influencing variables rather than relying on the correlation coefficient.

10.0 Conclusion

In this unit, the concept of correlation as well as how to compute correlation coefficient through several methods were adequately highlighted with many examples. Furthermore, the concepts of coefficient of determination and alienation became clearer to students. The computation of standard error of r and that of probable error of r were given attention in this unit. Conclusively, the fallacy of accepting that variable x caused variable y to happen was fully explained.

11.0 Summary

In the unit, students have been exposed to:

(i) the fact that correlation coefficient between two variables is just indicative of linear relation between two variables and not a causal/effect relationship..

(ii) how to compute coefficient of correlation with Pearson Product Moment Formula and that of Spearman Rank Order formula.

(iii) how to calculate coefficient of determination and alienation.

(iv) how to calculate standard error of r and probable error of r.

(v) understand the fallacy involved in the interpretation of the calculated correlation coefficients.

6.0 Tutor Marked Assignment

(1) Use raw score method to find the coefficient of correlation for the following 10 sample observations.

X	5	7	8	4	9	3	2	5	4	3
Y	2	4	5	5	6	5	4	4	3	2

(2) Compute the Probable error of r PE (r) if r = .72.

7.0 References/ Further Readings

Adewoye, S. O. (2004): Basic Statistics for Engineering, Economics and Management, Lagos: Olukayode Ojo Commercial Enterprises.

Clark, G.M. and Cooke, D. (2004) Basic Course in Statistics (5th ed) New York. Oxford University Press Inc.

Levin, R.I. and Rubin, D.S. (1994) Statistics for Management (7th ed) New Jersey: Prentice Hall Inc.

Monga, G.S. (2007). Mathematics and Statistics for Economics (Second Revised Edition) New Delhi, Vikas Publishing house PVT Ltd.

Salami, K.A. (1999) Descriptive Statistics for Beginners. Oyo.: Odumatt Press and Publishers.

Salami, K.A. (2001). Basic Statistics and Data Processing in Education. In Adeyanju A. (ed) Introduction to Educational Management, Oyo Green Light Press and Publishers.

Salami, K.A. (2001). Introduction to Basic Statistics in Management. In Adeyanju A. (ed) Introduction to Educational Management, Oyo Green Light Press and Publishers.

Unit 2 CHI SQUARE (X^2) TEST

CONTENT

1.0 INTRODUCTION

In this unit, students will be exposed to the fact that hypotheses about the goodness-of-fit of data to a discrete or continuous distribution can be examined using the X^2 statistic. Also the test of independence of row and column classifications in a two-dimensional or 3 x 4 table also uses X^2. Further more, students will discover from this unit that chi-square (X^2) statistic is a non-parametric testing method which does not depend on any assumption about the population distribution.

2.0 OBJECTIVES

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

(xiv) identify the uses of chi-square (X^2).statistics

(xv) to test the goodness-of-fit of any data.

(xvi) to test the independence of attributes .

3.0 Main Content

3.1 The Chi-Square (X^2) Test

The chi-square test is a non-parametric inferential statistical method commonly used in the analysis of frequencies or nominal data. As a nonparametric statistic, it makes no restrictive assumptions about the distribution of scores. Because of this, the method becomes useful in education and other behavioural sciences; particularly in the analysis of data in the form of frequencies or categories.

The chi-square is a two-tailed test. It can only indicate whether or not a set of observed frequencies differ significantly from the corresponding set of expected frequencies and not possibly the direction in which they differ.

3.3 The Uses of Chi-square Statistic (X^2)

X^2 distribution is used for the following tests:

(i) to test the goodness-of-fit, that is, to test the difference between theoretical and observed frequencies.

(ii) to test the independence of two variables.

(iii) to test if the population variance has a specified value.

3.4 The X^2 Goodness-of-fit Test

Checking the goodness-of-fit of X^2 is to inform us whether or not a set of observed frequencies fits closely to the theoretical or expected frequencies.

The general formula for this statistic is given as:

$$X^2 = \sum \frac{(O-E)^2}{E}$$ Where,

O = Observed frequency

E = Expected frequency

$\sum$ = Summation sign.

3.3.1 Exercise 2.1:

120 students were randomly selected from Federal Government College in Lagos with a view to determing their subject preference. From the records, it was observed that 65 preferred science while 55 preferred social sciences. With this information, you are requested to find X^2 statistic.

Table 2.1: Subject Preference of 120 Students

Subject Preference	
Science	Social Sciences
65	55

Table 2.2: Observed and Expected Frequencies of subject Preference of 120 students

Subject Preference		
	Science	Social Sciences
Observed f	65	55
Expected f	60	60

X^2 Calculation

Recall the formula as $X^2 = \sum \frac{(O-E)^2}{E}$

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

$$= \frac{(5)^2}{60} + \frac{(5)^2}{60}$$

$$= \frac{25}{60} + \frac{25}{60}$$

$$= 0.4167 + 0.4167$$

$$= 0.83 \text{ approximately.}$$

Self Assessment Exercise 1

Discuss the uses of chi-square statistic (X^2)

3.3.2 Necessary Conditions

Formulate null and alternative hypotheses to back the result. These hypotheses are:

Ho_1: There is no significant difference in the subject preference pattern of the students.

H_1: There is significant difference in the subject preference pattern of the students.

Here we choose 5% level of significance for this test while the degree of freedom here is n-1 = 1.

Decision to Make

Accept the null hypothesis if the calculated X^2 is less than the table value of X^2. And reject the null hypothesis if the calculated X^2 is greater than the table value.

3.3.3 Result

With this data, the calculated X^2 is 0.83 while the table value of X^2 is 3.84, df = 1 and $p<0.05$. Since the calculated value of X^2 is less than the table value of X^2 (0.83<3.84), we accept the null hypothesis that the difference observed in students preference for subject is by chance and not deliberate.

3.4 The X^2 Test of Independence

Another popular application of the X^2 test is in the testing for the independence of two variables. In this case, two factors, each having two or more levels are involved and the researcher wants to test whether or not the two variables are dependent or independent.

This type of information is usually presented in a contingency table. It is referred to as a contingency table because it displays data associated with two variables that are possibly contingent upon one another.

In a contingency table, it is normal to put the expected frequency in the same cell as the corresponding observed frequency. However, the expected frequencies are often differentiated by enclosing them in a bracket.

Exercise 2.2: To study the attitude of students to mathematics, 300 students from three socio-economic status were randomly selected from Federal Government School, Oyo They were given a questionnaire based on a 4 point Likert Scale. The result is as shown below.

Table 8.3: Attitude of students towards Mathematics

Socio-Economic Status	**Strongly Agreed**	**Agreed**	**Disagreed**	**Strongly Disagreed**	**Total**
High	20	15	30	15	80
Middle	40	35	15	10	100
Low	50	39	18	13	120
Total	**110**	**89**	**63**	**38**	**300**

With this information we want to know whether or not students attitude towards mathematics is dependent on the socio-economic status of parents.

3.4.1 Hypothesis Formulation

The necessary null hypothesis to be formulated under the X^2 test of independence is:

Ho: The attitude of students towards mathematics is significantly independent of socio-economic status of parents.

H_1: The attitude of students towards mathematics is significantly dependent on socio-economic status of parents.

3.4.2 Calculation of Expected Frequencies

The next operation is to calculate the expected frequencies based on the following formula.

$$E(RC) = \frac{r \times c}{N}$$

Where:

E(RC) = expected frequencies in the cell

r = total row frequency

c = total column frequency

N = total frequency involved.

3.4.3 The Calculation

Roll 1 cell 1: $E = \frac{80 \times 110}{300} = 29.33$

Roll 1 cell 2: $E = \frac{80 \times 89}{300} = 23.73$

Roll 1 cell 3: $E = \frac{80 \times 63}{300} = 16.80$

Roll 1 cell 4: $E = \frac{80 \times 38}{300} = 10.13$

Roll 2 cell 1: $E = \frac{100 \times 110}{300} = 36.67$

Roll 2 cell 2: $E = \frac{100 \times 89}{300} = 29.67$

Roll 2 cell 3: E = $\frac{100x63}{300}$ = 21.00

Roll 2 cell 4: E = $\frac{100x38}{300}$ = 12.67

Roll 3 cell 1: E = $\frac{120x110}{300}$ = 44.00

Roll 3 cell 2: E = $\frac{120x89}{300}$ = 35.60

Roll 3 cell 3: E = $\frac{120x63}{300}$ = 25.20

Roll 3 cell 4: E = $\frac{120x38}{300}$ = 15.20

These expected frequencies are now presented along with the corresponding observed frequencies in a 3 x 4 contingency table as shown below. The figures in brackets are the expected frequencies.

Table 8.2: A 3 x 4 contingency Table

Socio-Economic Status	**SA**	**A**	**D**	**SD**	**Total**
High	20 (29.33)	15 (23.73)	30 (16.80)	15 (10.13)	80
Middle	40 (36.67)	35 (29.67)	15 (21.00)	10 (12.67)	100
Low	50 (40.0)	39 (35.60)	18 (25.20)	13 (15.20)	120
Total	**110**	**89**	**63**	**38**	**300**

3.4.4 The computation is as follows:

$$X^2 = \frac{(20-29.33)^2}{29.33} + \frac{(15-23.73)^2}{23.73} + \frac{(30-16.80)^2}{16.80} + \frac{(15-10.13)^2}{10.13}$$

$$+ \frac{(40-36.67)^2}{36.67} + \frac{(35-29.67)^2}{29.67} + \frac{(15-21.00)^2}{21.00} + \frac{(10-12.67)^2}{12.67}$$

$$+ \frac{(50-44.00)^2}{44.00} + \frac{(39-35.60)^2}{35.60} + \frac{(18-25.20)^2}{25.20} + \frac{(13-15.20)^2}{15.20}$$

$X^2 = 2.97 + 3.21 + 10.37 + 2.34$

$+ 0.30 + 0.96 + 1.71 + 0.56$

$+ 0.82 + 0.32 + 2.06 + 0.32 = 25.94.$

The calculated $X^2 = 25.94.$

3.4.5 Table Value of X^2

To determine the table value of X^2, we need to determine the associated degree of freedom.

df = (R-1) (C-1)

R = the number of rows

C = the number of columns.

So in this case df = (3-1) (4-1)

df = 6

We shall also use 5% level of significance usually written as $p<0.05$.

Therefore, checking df = 6 and at $p<0.05$ level of significance on the X^2 table we get 12.59. So that

calculated $X^2 = 25.94$

table value of $X^2 = 1.64$.

3.4.5 Decision Criteria

Since the calculated value of X^2 is greater than the table value of X^2, (25.94>1.64) we reject the null hypothesis and accept the alternative hypothesis which says the attitude of students towards mathematics is dependent on socio-economic status of parents.

15.0 Conclusion

In this unit, students have been informed that chi-square test is a non-parametric test, which does not require or depend on any assumption about the population distribution. Furthermore, students were taught how to use chi-square to test for the goodness-of-fit of any data as well as test the independence of two variables. This unit also treated for students how to formulate hypothesis and how to compare the calculated X^2 with the table figure and that if calculated X^2 is greater than the table value of X^2 we should reject the null hypothesis but if it is less than the table value of X^2, we should accept the null hypothesis.

16.0 Summary

In the unit, students have learnt:

(vi) the uses of chi-square (X^2) as a nonparametric statistic.

(vii) how to test for the goodness-of-fit of any distribution .

(viii) how to test for the independence of two variables.

(ix) how to calculate the degree of freedom for the distribution.

(x) how to take decision on whether to accept or reject a null hypothesis.

6.0 Tutor Marked Assignment

i. To study the effect of gender on students academic performance, 50 boys and 50 girls were randomly selected from schools, colleges and universities. Table below illustrate this.

Table 8.1: 100 students selected on Gender basis from Institutions

SEX	SCHOOL	COLLEGE	UNIVERSITY	TOTAL
Boys	10	15	25	50
Girls	25	10	15	50
Total	**35**	**25**	**40**	**100**

With this information, find whether academic performance depends on sex (gender) given df = 2 and $p<0.05$ with critical x^2 as 5.99

REFERENCES

Adewoye, S. O. (2004): Basic Statistics for Engineering, Economics and Management, Lagos: Olukayode Ojo Commercial Enterprises.

Clark, G.M. and Cooke, D. (2004). Basic Course in Statistics (5th ed) New York. Oxford University Press Inc.

Levin, R.I. and Rubin, D.S. (1994). Statistics for Management (7th ed) New Jersey: Prentice Hall Inc.

Monga, G.S. (2007). Mathematics and Statistics for Economics (Second Revised Edition) New Delhi, Vikas Publishing house PVT Ltd.

Salami, K.A. (1999). Descriptive Statistics for Beginners. Oyo. Odumatt Press and Publishers.

Salami, K.A. (2001). Introduction to Basic Statistics in Management. In Adeyanju A. (ed) Introduction to Educational Management, Oyo Green Light Press and Publishers.

Salami, K.A. (2001). Basic Statistics and Data Processing in Education. In Adeyanju A. (ed) Introduction to Educational Management, Oyo Green Light Press and Publishers.

Unit 3 STUDENT S t STATISTIC

CONTENTS

1.0 Introduction

Students will recall that unit 1 treated correlation and we arrived at a correlation coefficient, which could indicate positive or negative relationship between variables. Unit 2 also treated chi-square with its statistic, which is basically used for opinion polls or perceptions of people about events or objects. This unit 3 will treat another statistical tool that is good for data processing in educational management. This statistical tool called student s t distribution is used for small sample of 30 or less and it was invented by a British W.S. Gosset (1876-1937), who worked as a statistician for Guinness, and writing under the pen name of student. The statistical tool became so popular among the researchers and educationists because of its attributes

2.0 Objectives

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

(i) identify the relationship between large and small samples.

(ii) appreciate the invention of student t-statistic.

(iii) know how to use student t-statistic.

(iv) learn from the application of t-statistic, the basic principle.

3.0 Main Content

3.1 Relationship between z and t Distributions

When testing the significance of a mean of normally distributed variables, we use the statistic

$$z = \frac{\bar{x} \quad \mu}{\sigma\sqrt{n}}$$

This distribution contains only one variable element in it, x which varies from sample to sample: here n is fixed and and are known. But in real world practice, is not usually determined or known and because of that, it has to be estimated from the sample using.

$$S^2 = \frac{1}{n\ 1} \sum_{i=1}^{n} (x \bar{x})^2$$

as the unbiased estimator of ². If the sample is large (n>30) s² is likely to be very close to². So we treat s² as if it is². But we cannot do this in small sample (n<30). In such a situation, s² has a high chance of being very different from². How then do we cope with this difficulty?

3.2 The Origin of t-distribution

W.S. Gosset, a British (1876-1937), who worked as a statistician for Guinness and writing under the pen name student , invented what is known today as student s t-distribution with the formula

$$\frac{\bar{x} \quad \mu}{\sigma\sqrt{n}}$$

which is usually denoted by t and represents a small sampling distribution. This distribution has two variable elements in it, $\bar{x}$ and S, both of which alter or change each time a new sample of size n is taken.

3.3 Small Sampling Theory

A study of sampling distribution of statistics for small samples (n<30) is called small sampling theory on exact sampling theory since the results obtained from small sampling theory hold for large as well as small samples.

3.4 Student t-formula

It is written as

$$t = \frac{x \quad \mu \quad \sqrt{n} - 1}{s}$$

where $\sqrt{\frac{n}{(N-1)s}}$

Example 3.1

A test of the breaking strength of 6 ropes manufactured by a local company in Oyo showed a mean breaking strength of 7750N and a standard deviation of 145N. Whereas, the manufacturer claimed a mean breaking strength of 8000N. We want to check whether the manufacturer s claim is right at 0.05 and 0.01 level of significance.

The null hypothesis could be stated as follows:

Ho: *There is no significant difference between the observed mean strength of 6 ropes and the manufacturer s claimed mean strength of the same ropes.*

(a) Solution

This is one tailed

Decision

(1) Reject Ho if t calculated is more than the t⟨ (tabulated)

(2) Accept Ho if t calculated is less than the t⟨ (tabulated)

Here the $\bar{x}$ = 7,750, ∝ = 8,000, df = N-1 = 6-1 = 5.

$$t = \frac{7750-8000}{145/\sqrt{5}} = -3.855$$

or = -3.85 approximately.

(b) Tabulated value of t at P < 0.05 with d.f. = 5

is 2.015

(c) First Conclusion

Since t calculated is -3.86 and t⟨ is -2.015 at p<0.05 with d.f = 5, we will reject the null hypothesis and conclude that the manufacturer s claim is true.

(d) Tabulated value of t at P < 0.01 with d.f = 5

is -3.365

(e) Second Conclusion

Since t calculated (-3.86>-3.365) at p<0.01with d.f = 5 is greater than the t-table, we will reject the null hypothesis and accept the manufacturer s claim as true.

Example 3.2 Test involving two different means

The I.Q s (Intelligence quotients) of 16 male students from Z.K.Y school showed a mean of 107 with standard deviation of 10, while the I,Q s of 14 female students from WXY school showed a mean of 112 and standard deviation of 8. We want to check whether there is any difference in the two means.

This is a case of two random samples of sizes N_1 and N_2 drawn from normal populations whose standard deviations are equal S_1 and S_2 and means $_1$ and $_2$ $\bar{x}$

Null and alternative hypotheses to be formulated are:

Null hypothesis: There is no significant difference between the IQ mean of male and that of female students.

Alternative Hypothesis: There is significant difference between the IQ mean of male and that of female students.

Self Assessment Exercise 1

Mention the originator and inventor of Student t statistic. What is it formula?

Solution

Then,

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sigma\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

where $\sigma = \sqrt{\frac{N_1S_1^2 + N_2S_2^2}{N_1 + N_2 - 2}}$

and the degree of freedom d.f. is $N_1 + N_2 - 2$

(b) Decision

(i) Reject Ho if t-calculated is greater than t-table.

(ii) Accept Ho if t-calculated is less than t-table.

(c) **Calculated value of** σ

$$\sigma = \sqrt{\frac{N_1S_1^2 + N_2S_2^2}{N_1 + N_2 \quad 2}}$$

$$= \sqrt{\frac{16(10)^2 + 14(8)^2}{16 + 14 - 2}}$$

$$= \sqrt{\frac{1600 + 896}{16 + 14 \quad 2}}$$

$$= \sqrt{\frac{2496}{28}}$$

$$= \sqrt{89.14}$$

r = 9.44

with r as 9.44, we can now compute t

$$t = \frac{107-112}{9.44\sqrt{\frac{1}{16} + \frac{1}{14}}}$$

$$= \frac{-5}{9.44 \rightarrow 0.1339} = -1.447$$

t = -1.45 approximately

t⟨ at p<0.05 with d.f = 28 = -2.048

or -2.05 approximately

(d) First Conclusion

Since t calculated -1.45 is less than t⟨ -2.048 i.e. (-1.45<-2.048) we will accept the null hypothesis and conclude that there is no significant difference in the I Q s of the two groups.

(e) t⟨ at 0.01 with d.f = 28 is 2.763

Therefore the interval t⟨ of 0.01 is between -2.763 and 2.763.

(d) Second Conclusion

Since t calculated -1.45 is less than the t-table 2.762, we shall accept the null hypothesis as true and conclude that there is no statistically significant difference between the I Q s of the two groups.

3.5 Application of student t test Statistic: A Case Study

3.5.1 To demonstrate the application of student t test, a hypothetical study is hereby presented below.

Title

Socio-Economic Factors As Determinants of Career in Teaching Profession in South-West Nigeria:

Population and Sample

The population of the study includes all male and female teachers in South-West Nigeria. Out of this, two hundred (200) teachers 100 males and 100 females were randomly selected from twenty schools from South-West states. Both random and purposive sampling techniques were used to select the sample.

Instrumentation

The instrument used for the study was the Teaching Profession Preference Scale (TPPS). It is a twenty-items questionnaire that adopts the four Point Likert scale format.

Validity of Instrument

The instrument after construction was given to experts in the field to critique. Their various suggestions were incorporated into the final draft of the questionnaire. This ensured the face validity of the instrument.

Reliability of Instrument

The test-retest reliability method was used to test the reliability of the instrument. The reliability coefficient was found to be 0.81 which was assumed to be acceptable.

Data Administration

The instrument was administered in the selected schools through the help of some teachers.

Data Analysis

The data collected were analyzed using among others t-test statistical method.

Hypotheses Formulation

Two null hypotheses were formulated and tested at $P<0.05$ level of significance.

> ***Ho_1:*** *There is no significant difference between the perceptions of Male and female teachers on the socio-economic factors as distractor in teaching profession.*
>
> ***Ho_2:*** *There is no significant difference between the perceptions of young and old teachers on the socio-economic factors as distractor in teaching profession.*

The Results

Hypothesis 1: There is no significant difference between the perception of male and female teachers on the socio-economic factors as distractor in teaching profession.

Table 3.1: t-test analysis of teachers perceptions by gender on socio-economic factors as Teaching profession distractor.

Gender	**No**	$\bar{x}$	**SD**	**df**	**Cal. t**	**Table t.**	**P**
Male	100	48.65	10.38	198	5.96	1.96	S
Female	100	40.25	9.64				

The calculated t-value was 5.96 at P<0.05 and with df = 198. This value was found to be greater than the table value of t which was 1.96 (i.e. 5.96 > 1.96). This indicates that there is significant difference between male and female teachers perceptions on socio-economic factors as distractor in teaching profession. By this result, the null hypothesis was thus rejected.

Hypothesis 2: There is no significant difference between the perceptions of young and old teachers on the socio-economic factors as distractor in teaching profession.

Table 3.2: t-test analysis of teachers perceptions by age on socio-economic factors as distractor in teaching profession.

Status	**No**	$\bar{x}$	**SD**	**Df**	**Cal. t**	**Table t.**	**P**
Young	100	45.92	11.48	198	5.91	1.96	S
Old	100	37.23	9.35				

The calculated t-value of 5.91 at P<0.05 level of significance and with df = 198 was found to be greater than the critical/table value of t which is 1.96. (5.91 > 1.96).

This indicates that significant difference exists between the perceptions of the young and old on the socio-economic factors as distractor in teaching profession. The null hypothesis was therefore rejected.

3.5.2 How did we arrive at t-calculated of 5.96 and 5.91 respectively?

Using the following formula

$$t = \frac{\bar{x}_1 \quad \bar{x}}{\sigma\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

Where $\sigma = \sqrt{\frac{N_1 S_1^2 + N_2 S_2^2}{N_1 + N_2 \quad 2}}$

the degree of freedom df = $N_1 + N_2 \quad 2$

Hypothesis 1. $\bar{x}_1 = 48.65, \bar{x}_2 = 40.25$, $SD_1 = 10.38$, $SD_2 = 9.64$

Calculated value of σ is

$$= \sqrt{\frac{100\ (10.38)_2 + 100\ (9.64)^2}{100 + 100 \quad 2}}$$

$$= \sqrt{\frac{100\ (10.38)_2 + 100\ (9.64)^2}{198}}$$

$$= \frac{\sqrt{100\ (107.74) + 100\ (92.93)}}{198}$$

$$= \frac{\sqrt{10774 + 9293}}{198}$$

$$= \frac{\sqrt{20067}}{198}$$

$$= \sqrt{101.34}$$

$$= 10.07$$

$$t = \frac{48.65 \quad 40.25}{10.07\sqrt{\frac{1}{100} + \frac{1}{100}}}$$

$$t = \frac{8.4}{10.07\sqrt{(0.02)}}$$

$$= \frac{8.4}{10.07\ (0.14)}$$

$$= \frac{8.4}{1.14}$$

$$t = 5.96$$

Hypothesis 2

Using the same formula for hypothesis 2, we have

$\bar{x}_1 = 45.92 \quad \bar{x} = 37.23$, $SD_1 = 11.48$, $SD_2 = 9.35$

calculated value of σ

$$\sigma = \frac{\sqrt{100\ (11.48)_2 + 100\ (9.35)^2}}{100 + 100 \quad 2}$$

$$= \frac{\sqrt{100\ (131.79) + 100\ (87.42)}}{198}$$

$$= \frac{\sqrt{13179 + 8742}}{198}$$

$$= \frac{\sqrt{21921}}{198}$$

$= \sqrt{110.71}$

$\sigma = 10.52$

$$t = \frac{45.92 \quad 37.23}{10.52\sqrt{\frac{1}{100} + \frac{1}{100}}}$$

$$= \frac{8.69}{10.52\sqrt{(0.02)}}$$

$$t = \frac{8.69}{10.52\ (0.14)}$$

$$= \frac{8.69}{1.47}$$

$t = 5.91$

10.0 Conclusion

This unit has exposed the students to:

(i) the relationship between z (large sample size) and t (small sample size) in hypothesis testing,

(ii) the origin of student t-distribution,

(iii) the student t-distribution formula

(iv) many applications of student t statistic.

11.0 Summary

In this unit, students have been taught:

(i) the relationship between z and t distributions

(ii) the origin of student t distribution

(iii) how to use student t distribution through many examples

(iv) how to formulate null and alternative hypotheses

(v) that student t distribution is concerned with small samples size i.e. (n<30).

6.0 Tutor Marked Assignment

1. Giving two independent means and standard deviation for 2 sample sizes of 80 and 47 as:

$\bar{X}_1 = 10.25$, $\bar{X}_2 = 6.44$, $SD_1^2 = 7.12$, $SD_2^2 = 4.11$

calculate the t-statistic, given the critical t as 1.96 at $p<0.05$ level of significance with df = 125. Indicate whether we should accept or reject the formulated null hypothesis, which says there is no significant difference in the two means.

7.0 References/Further Readings

Adewoye, S. O. (2004): Basic Statistics for Engineering, Economics and Management, Lagos: Olukayode Ojo Commercial Enterprises.

Hildebrand, D.K. & Lyman, O. (1998). Statistical thinking for managers (4th Ed.) California: Duxbury Press.

Levin, R.I. &Rubin, D.S. (1997). Statistics for Management (7th Ed.) New Jersey: Prentice Hall International.

Monga, G.S. (2007). Mathematics and Statistics for Economics (Second Revised Edition) New Delhi. VIKAS Publishing House PVT Ltd.

Salami, K.A. (1999). Descriptive Statistics for Beginners. Oyo Odumatt Press and Publishers.

Salami, K.A. (2001). Introduction to Basic Statistics in Management. In Adeyanju, A. (Ed.) Introduction to Educational Management. Oyo Green Light Press and Publishers.

Unit 4 THE REGRESSION ANALYSIS

CONTENTS

1.0 Introduction

Every day education managers make personal and professional decisions that are based upon predictions of future events. To make these forecasts, the education managers rely upon relationship between what is already known and what is to be estimated. If managers can determine how the known variable is related to the future event, then, such exercise will help decision making process considerably. The subject of this unit therefore is how to determine the relationship between variables using regression analysis.

2.0 Objectives

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

(i) identify the relationship between prediction and regression

(ii) identify both the Simple and Multiple Regression Equations

(iii) familiarize with the method for fitting the regression line

(iv) compute simple regression line equation

(v) work out multiple regression equations.

3.0 Main Content

3.1 Prediction and Regression Defined

A prediction is a guess about the value of an observations that is to be drawn from a population

- If we know nothing about the population, then, any prediction arising from such observation becomes a blind prediction. Therefore some knowledge of the population is very necessary so as to make meaningful predictions
- A prediction variable is one about which predictions are made
- A predictor variable is one that provides relevant information for prediction
- The technique of using one variable to make prediction about another is called regression analysis. In regression analysis, we make use of known values of one variable to predict unknown value of the other
- Regression is the estimation of unknown values or the prediction of one variable from known values of other variables.

3.2 Regression Equation

The linear regression relationship between Y and X can be written in the form of straight line equation

$$Y = a + bx$$

Where a, b, are constants determining the position of line b being its slope and a the intercept on the Y- axis.

3.3 Method of Fitting the Regression Line

There are two methods commonly used for fitting the regression line. One is unscientific while the other is scientific and is acceptable in research work.

3.3.1 Freehand Method

The first method is called freehand method, which is unscientific in the sense that fitting a line between two poles or pair of points could be subjective if done with freehand without a guide. More so if many people are involved, the fitting lines may not be uniform.

3.3.2 The Method of Least Square (Scientific method)

The function of the least square method is to minimize the error between the estimated points on the line and the actual observed points that were used to draw it.

When this is done, statistician called it a good fit of the lines. The equation for the estimated line is

$$\hat{Y} = a + bx \qquad . (1)$$

$\hat{Y}$ symbolizes the individual values of the estimated points, that is, those points that lie on the estimated line.

In equation (1) above ( ) is the dependent variable while (X) is the independent variable and (a) and (b) are the unknown constants. The major function of regression analysis is to determine the values of these constants. There are two commonly used methods for calculating variables (a) and (b). The first one is called the Raw Score method, while the second method is known as the Deviation method.

3.4.1 The Raw Score Method

The equation for determining a and b is given as;

$$a = \frac{(Y X^2)(XY)(X)}{N X^2 (X)^2}$$

$$b = \frac{N XY (X)(Y)}{N X^2 (X)^2}$$

Where

a = intercept of Y

b = slope of the line

3.4.2 The Deviation Method

$$Y- C + MX$$

$$M = \frac{(X\bar{X})(Y-\bar{Y})}{(X-X)^2}$$

$$C = \bar{Y} - M\bar{X}$$

Where M is the slope of the line

C is the intercept of Y

Alternatively, we can use this formula;

$$b = \frac{XY \quad nXY}{X^2 \quad NX^2}$$

And $a = \bar{Y} \quad b\bar{X}$

Exercise 4.1

The commissioner of education in Oyo state is interested in the relationship between the age of some schools and the annual maintenance cost spent on them. To determine this relationship, four schools with 6,5,3 and 2 years old respectively were randomly selected and whose maintenance costs were N900, N800, N600 and N500 respectively per year. For the four schools, the commissioner is highly interested in knowing the regression equation to be used for forecasting future expenses based on the age of the schools.

Solution

The first step in calculating the regression line for this problem is to re-organize the data as outlined below:

Table 4.1: Regression Analysis for School buildings

Schools	AGE(x)	Maintenance cost (Y) (N00)	XY	X^2
A	6	9	54	36
B	5	8	40	25
C	3	6	18	09
D	2	5	10	04
	16	28	122	74

Mean = $\bar{X} = \frac{16}{4} = 4$

Mean = $\bar{Y} = \frac{28}{4} = 7$

Using the Raw Score Method, a and b become: (refer to the equation)

$$a = \frac{28(74) \quad (122)\,(16)}{4(74) \quad (16)^2}$$

$$= \frac{2072 \quad 1952}{296 \quad 256}$$

$$= \frac{120}{40}$$

$$= 3$$

And

$$b = \frac{4(122) - (16)(28)}{4(74) - (16)^2}$$

$$= \frac{488 - 448}{296 - 256}$$

$$= \frac{40}{40}$$

$$= 1$$

The regression equation becomes Y = 3 + 1x. With this equation, a school building built 12 years ago would need N1,500 as maintenance expenses.

e.g. Y = 3 + 1(12)

= 3 + 12

= N1,500

Therefore, as (X) which is the age of the school building changes, so will the maintenance cost or expenses increase..

3.4.3 The Deviation Method

We can also use the deviation method to arrive at the same answer.

Solution

Table 4.2:Regression analysis for school building

School	Age (x)	Expenses (y)	$(X-\bar{X})$	$(Y-\bar{Y})$	XY	$(X-\bar{X})2$
A	6	9	2	2	4	4
B	5	8	1	1	1	1
C	3	6	-1	-1	1	1
D	2	5	-2	-2	4	4
	16	28			10	10

$\bar{X} = \frac{16}{4} = 4$

4

$\bar{Y}$ = 28 = 7
4

Here $(X-\bar{X})(Y-\bar{Y})$ = 10

$(X-X)^2$ = 10

Substituting all these values into the equations will give;

$M = \frac{10}{10} = 1$

C = 7- 1(4)

= 7- 4

= 3

So that Y = 3 + 1x, the same equation line as the Raw Score Method.

Self Assessment Exercise 1

Find the regression of Y on X from the following data

X 6, 5 , 3 ,2 ,4

Y 9, 8, 6, 5, 3

3.5 Graphical illustration of the regression line

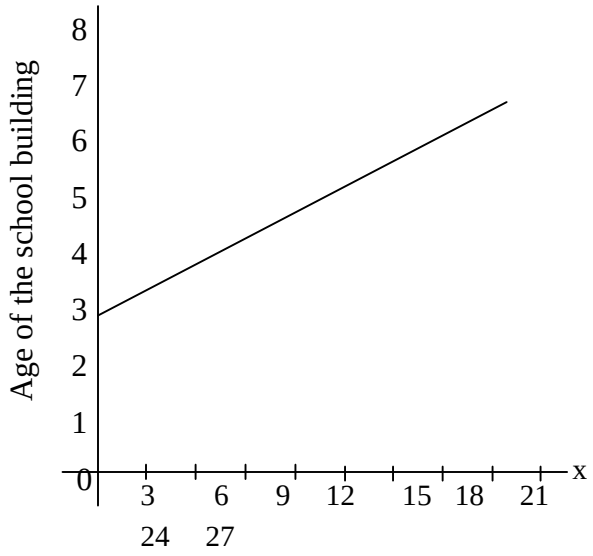


Fig. 4.1: Maintenance expenses ((in hundreds of naira)

3.6 Uses of regression analysis

Education managers can make use of this linear regression analysis for decision making, not only for building maintenance expenses but for many of the school facilities, equipments and infrastructure. After all, the bane of Nigeria educational system is lack of maintenance culture as well as accurate data base. It is on this premise that this type of statistical analysis becomes a must and necessity for all education managers.

3.7 Multiple Regression Analysis

Unlike the simple regression analysis, multiple regression analysis involves more than one variable. It is capable of taking two or more variables. Multiple regression equation is written as;

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 \quad .. + bnXn.$$

Where Y is the estimated value of Y

a is the intercept of Y

X_1, X_2, X_3 are the values of independent variables

b_1, b_2, b_3 are the slopes associated with X_1, X_2, X_3 respectively.

In this unit, we shall experiment with only two independent variables. Although, the same principle is applicable to any number of independent variables of our desire.

A principal of a private school is trying to estimates the monthly amount of unpaid school fee discovered by his cashier. In the past, the principal usually estimate this figure on the basis of what the class teachers collect for him. In recent years, however, the class teachers collections have become an erratic predictor of the actual unpaid school fees. As a result, the principal requested a statistician to prepare a statistical tool which he can use for accurate estimation of these unpaid school fees. He therefore presents the following previous ten months collection to the statistician.

Table 4.3 Data on school fee collection

Month	Class teacher collection (X_1)	Cashier collection (X_2)	Actual unpaid school fee discovered
January	45	16	29
February	42	14	24
March	44	15	27
April	45	13	25
May	43	13	26
June	46	14	28
July	44	16	30
August	45	16	28
September	44	15	28
October	43	15	s27

The next step is to re-arrange this data in such a way that will fit the required equation as shown below

Table 4.4: Re-arrangement of data in tables 9.3

Y	X_1	X_2	X_1Y	X_2Y	X_1X_2	X_{12}	X_{22}	Y^2
29	45	16	1,305	464	720	2,025	256	841
24	42	14	1,008	336	588	1,764	196	576
27	44	15	1,188	405	660	1,936	225	729
25	45	13	1,125	325	585	2,025	169	625
26	43	13	1,118	338	559	1,849	169	676
28	46	14	1,288	372	644	2,116	196	784
30	44	16	1,320	480	704	1,936	256	900
28	45	16	1,260	448	720	2,025	256	784
28	44	15	1,232	420	660	1,936	225	784
27	43	15	1,161	405	645	1,849	225	729
272	441	147	72,005	4,013	6,485	19,461	2,173	7,428

With this information where

N = 10, $\bar{Y}$ = 27.2, $\bar{X}_1$ = 44.1, $\bar{X}_2$ = 14.7

The required multiple regression equation is

$Y = na + b_1X_1 + b2\ X2$

$XiY = a\ X_1 + bi\ X_{12} + b2\ X_1\ X_2$

$X2Y = a\ X_2 + b_1X_1\ X_2 + b_2X_{22}$

The above information gives us three equations with three unknown constants (a, b_1, b_2). Substituting the figures in table 9.4 into the equation, we get

$272 = 10a + 441b_1 + 147b_2$ (i)

$12{,}005 = 442a + 19461b_1 + 6{,}485b_2$ (ii)

$4{,}013 = 147a + 6{,}485b_1 + 2{,}173b_2$ (iii)

We can now find the values for the three constants (a, b_1, b_2) by solving the three equations simultaneously as illustrated below

Step 1 multiply equ. 1 by -441

multiple equ. 2 by 10

and add 1 and 2 to eliminate constant a

$$- \quad 119{,}952 = 441a \quad 194{,}481b_1 \quad - \quad 64{,}827b_2$$

$$\underline{120{,}050 = 441a + 194{,}610b_1 + 64{,}850b_2}$$

$$98 = 129b_1 + 23b_2 \qquad \text{(iv)}$$

Step 2 multiply equ. 1 by -147

Multiple equ. 3 by 10

And add equ. 1 and 3 to eliminate constant a

$$-39{,}984 = 147a - 64{,}827b_1 - 21{,}609b_2$$

$$\underline{+40{,}130 = 147a + 64{,}850b_1 + 21{,}730b_2}$$

$$146 + 23bi + 121b2 \qquad \text{(v)}$$

Step 3 multiply equ. 4 by - 23

Multiply equ. 5 by 129

And add equ. 4 and 5 to eliminate bi

$$- \quad 2{,}254 = - 2{,}967b_1 - 5{,}29b_2$$

$$\underline{+ \quad 18{,}834 = +2967b_1 + 15609b_2}$$

$$16{,}580 = 15{,}080b_2$$

We can now determine the value of b_2 with the result in step 3, as follows:

$$16580 = 15080b_2$$

$$b_2 = \frac{16580}{15080}$$

$$b_2 = 1.099$$

Step 4 To find the value of b_1, substitute the value of b_2 into equation 4

$$98 = 129b_1 + 23b_2$$

$$98 = 129b_1 + 23x1.099$$

$$98 = 129 + 25.227$$

$$98 - 25.227 = 129b_1$$

$$72.772 = 120b_1$$

$$b_1 = \frac{72.772}{129}$$

$$b_1 = 0.564$$

Step 5: Substitute the value of b_1 and b_2 into equation 1

$$272 = 10a + 441b_1 + 147b_2$$

$$272 = 10a + 441(0.564) + 147(1.099)$$

$$272 = 10a + 248.72 + 148.32$$

$$125.4 = 10a$$

$$a = \frac{125.4}{10}$$

$$a = 12.54$$

Step 6: Substitute the values of a, b_1, b_2 into the general two variables regression equation

$$Y = -12.54 + 0.564X_1 + 1.099X_2$$

This regression equation describes the relationship among the class teachers and the cashier s performance in school fees collections. The Principal can now use this equation monthly to estimate the amount of unpaid school fees. Again he can use it to forecast the unpaid school fees every month. In like manner the education manager can also use this regression equation to predict several variables in his establishment.

4.0 Conclusion

As a continuation of treatment of data processing in educational management, this unit has further treated another of the statistical tools called regression analysis. In doing that, prediction and regression concepts were defined. Regression equations, methods of fitting regression lines, the scientific method of least square, graphical illustration and the uses of regression analysis in education were highlighted

5.0 Summary

In this unit, you have learnt:

(i) the relationship between prediction and regression

(ii) how to use regression equations and methods of fitting the regression lines

(iii) how to fit the regression lines for both simple and multiple regression equations

(iv) the usefulness of regression analysis in educational management, and

(v) how education managers can use it for decision making

6.0 Tutor Marked Assignment

1. Find the regression of Y on X from the following data by using the Raw score method

X	5,6,7,8,9
Y	2,4,6,7,11

2. Below is 6 years turnover and profit before tax of AZK Manufacturing Co. Ltd.

Year	1991	1992	1993	1994	1995	1996
Turnover (in million)	106	125	147	167	187	220
Profit Before Tax (in million)	10	12	16	17	18	22

Action Required

i. plot the scatter diagram showing the relationship between profit before taxation and turnover.

ii. calculate the least squares regression line of profit before taxation on turnover.

iii. comment generally on your results.

7.0 References/ Further Readings

Adewoye, S. O. (2004): Basic Statistics for Engineering, Economics and Management, Lagos: Olukayode Ojo Commercial Enterprises.

Hildebrand, D.K. & Lyman, O. (1998). Statistical thinking for managers (4th Ed.) California: Duxbury Press.

Levin, R.I. &Rubin, D.S. (1997). Statistics for Management (7th Ed.) New Jersey: Prentice Hall International.

Monga, G.S. (2007). Mathematics and Statistics for Economics (Second Revised Edition) New Delhi. VIKAS Publishing House PVT Ltd.

Salami, K.A. (1999). Descriptive Statistics for Beginners. Oyo Odumatt Press and Publishers.

Salami, K.A. (2001). Introduction to Basic Statistics in Management. In Adeyanju, A. (Ed.) Introduction to Educational Management. Oyo Green Light Press and Publishers.