NATIONAL OPEN UNIVERSITY OF NIGERIA

SCHOOL OF SCIENCE AND TECHNOLOGY

COURSE CODE: HEM 707

COURSE TITLE: DISEASE ILLNESS AND SOCIETY

HEM 707:   **BIOSTATISTICS – 2CU**

COURSE GUIDE

NATIONAL OPEN UNIVERSITY OF NIGERIA

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

Turn to unit 1 and read the introduction and the objectives for the unit.

Assemble the study materials. Information about what you need for a unit is given in the Table of Contents at the beginning of each unit. You will almost always need both the study unit you are working on and one of the materials for further reading on your desk at the same time.

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 other sources. Use the unit to guide your readings.

Keep in mind that you will learn a lot by doing all your assignments carefully. They have been designed to help you meet the objectives of the course and will help you pass the examination. Submit all assignments as at when due.

Review the objectives for each study unit to confirm that you have achieved them. If you are not certain about any of the objectives, review the study material or consult your tutor.

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

When you have submitted an assignment to your tutor for marking, do not wait for its return before starting on the next unit. Keep to your schedule. When the assignment is returned, pay particular attention to your tutor's comments both on the tutor's-marked assignment form and also comments written on the assignment. Consult your tutor as soon as possible if you have any questions or problems.

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

*** **Tutors and tutorials**

There are 15 hours of tutorials provided in support of this course. You will be notified of the dates, times and location of these tutorials, together with the name and phone number of your tutor as soon as you are allocated a tutorial group.

Your tutor will mark and comment on your assignment, 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 tutor well before the due date. At least two working days are required for this purpose. They will be marked by your tutor and returned to you as soon as possible.

Do not hesitate to contact your tutor by telephone, e-mail or discussion board if your need help. The following might be circumstances in which you would find help necessary, contact your tutor 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 questions or problems with an assignment, with your tutor's comments on an assignment or with the grading of an assignment.

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

*** **Summary**

Heem605 intends to introduce biostatistics to you. Upon completing this course, you will be equipped with the basic knowledge of biostatistics, sources and use of statistics, types of survey and sampling techniques, data collection and analysis, data presentation and report writing. You will be able to answer the following questions.

1.0   What do we understand by the concepts of biostatistics?

2.0   What are the measures of Location?

3.0   How can we explain measures of variability?

4.0   What are the sources and of use of statistics?

5.0   How can we explain variable and measurements?

6.0   What are the measures of diseases?

7.0 What do we understand about the methods of data collection and analysis?

8.0 How can we explain data presentation?

9.0 What are the various stages of report writing?

HEM 707:   BIOSTATISTICS – 2CU

COURSE DEVELOPMENT

Course Developer
Prof. I. O. Orubuloye

Unit Writer

Programme Leader

Course Coordinator

NATIONAL OPEN UNIVERSITY OF NIGERIA

UNIT1

BIOSTATISTICS: BASIC CONCEPTS OF BIOSTATISTICS

Table of Contents

Introduction

After reading through the course guide, you will have a general understanding of what this Unit is all about and how it fits into the structure of this course. This is an important Unit in the programme in Diploma in HIV/AIDS Education and Management. The Unit focuses on the definition of biostatistics and types of statistics.

Objectives

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

Define the concept of biostatistics

Explain why study statistics

Identify types of statistics

Definition of Biostatistics

The word statistics has two meanings for social scientists or epidemiologist. First, it refers to any collection of numerical observations on an item or aggregates such as the number of students in the Open University, number of people who vote in a particular election or number of people suffering from a particular disease and so on. Second, statistics refers to a specialized field of study like mathematics or sociology. Statistics involves numbers, an understanding of it requires some basic knowledge of mathematics. But as social scientists or epidemiologists, we are more interested in the application of statistics as a tool.

The concept statistics is a more generic term while the concept biostatistics is the statistics relating to living things e.g.

Attendance at the STD clinic over a period of time.

The age distribution of deaths from HIV/AIDS over a period of time.

The number of people treated for malaria in a health centre in a month or one year.

Since human beings are different from each other, the information collected on two or more people is likely to be different. For example the number of people attending STD clinic in a particular health center are not likely to be of the same age, sex, marital status, number of partners and social background. So also the number of people treated or died from malaria are likely to vary from one clinic to another, from one place to another and from one year to another.

All these descriptions suggest that measurements or observations do not necessarily have the same magnitude and so it is subject to a lot of variation. This variation is central to the subject matter of statistics. Statistics as a discipline can therefore be defined as the "scientific methods" for the collection, summarization, presentation, analysis and interpretation of data.

Why Study Statistics

We study statistics for two main reasons. First, we are expose to a wide range of information in our lives daily. For example we are told

the number of candidates who sit for the Joint Admission Matriculation board Examination every year, the number who met the minimum requirements for admission and the number that are finally placed in the different universities. We are told about the number of motor accidents on our roads, the number of people who died from such accidents and number who suffer from injuries. We also read about the results of the distribution of certain products over a period of time. For example it is now customary for the NNPC to publish daily distribution of petroleum products across the country. These information may represent the truth. It may also be untrue. Statistics can be used to support almost anything. It is also easy to lie with statistics and use it purposely to distort the truth. Nevertheless, it is important to study statistics so that we can know how to evaluate published data, when to believe them, when to be skeptical and when to reject them.

The second reason for studying statistics is to secure employment in both private and public sectors of the economy knowledge of statistics is relevant for the understanding of nearly all aspect of daily lives. An employee may be asked to record or estimate the frequency of occurrence of certain events. For example, the number of students who sat or who are likely to sit for a particular examination or the number of people who are likely to purchase a particular product or the number of people who are likely to purchase a particular product or the number of people who are likely to vote for a particular political party in a national election. We may also want to know the characteristics of these people, the age, sex and other social background.

Although statistics is a separate discipline, it is essential to the social, biological and physical sciences because all scientists use observations of natural phenomena, through sample surveys or experimentation to gather facts, test hypotheses and develop theories. Statistics has become one of the major tools of the scientific method and is necessary for the development of knowledge in all fields of science.

3.2   Types of Statistics

There are two types of statistics, which are important for application in the social sciences. They are descriptive statistics and inferential statistics.
What is descriptive statistics? Descriptive Statistics provides the social scientists with graphical and numerical techniques for

describing or summarizing in a concise from complex and massive information.   For example, if 700,000 candidates applied for admission in to the Open and distance learning Education programme in a particular year, we might wan to know the characteristics of the candidates of the applicants.   We might wan to know their social background, their age, sex and their state of origin or their local government areas, descriptive statistics provides us the techniques of summarizing such mass of data.

Descriptive statistics can also be used to describe the characteristics of the people treated for malaria in a clinic over a period of time or women that attending ante-natal clinic. Descriptive statistics can also be used to make comparisons between two characteristics measured on every person in a group or between groups using the same characteristics. We might wan to examine the relationship between UME score and performance of students or class of degrees on graduation or the socioeconomic backgrounds of male and females at the time of their admission.

Descriptive statistics is important in preliminary analysis of data prior to the use of more rigorous statistical analysis.

What is inferential statistics?   Most social scientists study large groups in their investigations and it is often impossible to reach every member of the population. What social scientists then do is to select a fraction or a sample of the large population. Let us go back to the example of the 700,000 applicants for admission in the Open University programme. Suppose we wish to estimate the proportion of all the 700,000 applicants who are qualified for admission and who are willing to take up the offer if given admission. One way of doing this is to contact all the 700,000 applicants. It will be time consuming and almost impossible to visit and interview all the 700,000 candidates.   An easier and more efficient approach is to select a fraction or a sample of them. We can do a random sample of 2,500 of them or even less depending on the resources available for the exercise. We can then collect such information from the sample and use it to estimate for the total population.

Inferential statistics are useful in predicting election results.   For example, we might want to predict the outcome of local government election or presidential election, it is also not possible to contact all eligible voters even in a local government area, a sample of the registered voters is then taken and questions posed to them to know their voting behaviour.   The results are then analyzed to predict

possible outcome of the elections. Survey techniques are widely used in opinion surveys today and inferential statistics is the man tool for handling such information.

4.0 Conclusion

Statistics is an important tool for the social scientists. It has dual meaning for social researchers. It refers to numbers such as rates of deaths, number of applicants to the Open University and number of students per course. Statistics is also a distinct field of study. Statistics is essential to the social, biological and physical sciences because all scientists make use of observations of natural phenomena through surveys or experimentation, to gather facts, test hypotheses and develop theories. Statistics help social scientists in the understanding, objectivity and predictions of social phenomenon.

5.0 Summary

In this Unit, students have learnt about the definition of statistics in general, the distinction between general statistics and biostatistics. We have also discussed why we study statistics, the two types often used in social sciences and use of statistics.

6.0 Exercise

1. Define statistics and biostatistics.
2. Distinguish between descriptive and inferential statistics.
3. Discuss why we study and use of statistics.

7.0 Further Readings and Other resources

Blalock, H.M. Social Statistics, 2nd ed. New York: McGraw-Hill, 1972

7.0 Further Readings and Other Resources

I.O ORUBULOYE AND FOLAKEMI Oguntimehin. The study of Human Populations.
Centre of Population and health Research, Ado-Ekiti, Nigeria, 2000.

UNIT 2

BIOSTATISTICS: VARIABLES AND MEASUREMENT

Table of contents

1.0 Introduction

This Unit is a follow-up to Unit 1 which focused on the definition of biostatistics, types of statistics and use of statistics. This Unit will discuss definition of variables and examine the different scale of measurement: nominal, ordinal, interval and ratio.

2.0 Objectives

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

Define the concept variables
Examine the nature of measurement
Identify the different types of scales of measurement

3.0 Definition of variables

In Unit 1 we emphasized that social scientists like other scientists make use of observations of natural phenomena. Through surveys or experimentation, to gather facts, test hypotheses and develop theories.

When such observations on the same phenomenon remain constant in successive trials, the phenomenon is called constant and when they vary from trial to trial, the phenomenon is called a variable. In general, two types of variables are distinguishable in statistics. These are quantitative variable whose observation vary in magnitude from trial to trial or has numerical values and qualitative variables whose observations vary in kind but not in degree or are mainly categorical with no notion of numerical strength Examples of quantitative variables are age, size of the class of the Diploma in HIV Education and Management or size of the student population of the Open University. Qualitative variables on the other hand can be sex, education, religion and ethnic group. Quantitative variables can further be classified into discrete variable and continuous variable. When observations on a quantitative variable can assume only a countable number of values or whose values are distinct and separated or take whole numbers the variable is called a discrete variable. Examples of discrete variables are parity and age last birthday. When observations on a quantitative variable can assume any one of the countless numbers of values in a line interval or the value can assume decimal and fractions it is called a continuous variable. Examples of continuous variables are height, weight and exact age.

3.1 Measurement of Variables

It is generally assumed that human social behaviour is so complex and elusive that measurement is meaningless or impossible. Measurement always take place in a more or less complex situation in which innumerable factors may affect both the characteristics being measured and the process of measurement itself.

According to Selltiz, Jahoda, Deutsh and Cook, the basic problem in evaluating the results of any measurement is that of defining what shall be considered as variations due to error in measurement. They went further to consider some possible sources of differences in scores among a group of individuals. They are:

Differences due to transient personal factors

Differences due to statistical factors.
True differences in the characteristic, which one is attempting to
measure
True differences in other relatively stable characteristics of the
individual which affect his score.
Differences due to variations in administration.
Differences due to sampling of items.
Differences due to lack of clarity of the measuring instrument
Differences due to mechanical factors.
Differences due to factors in the analysis (processes of scoring,
coding, and computation.)

3.2   Scales of Measurement of variables

There are four scales on which variables are measured:

Norminal Scale: A norminal scale classifies qualitative objects into
categories by name.   This is the simplest scale and it is mainly
classificatory. It has no notion of numerical magnitude, as categories
are mutually exclusive and unordered. Examples of norminal scale
are ethnic group, sex, religion, marital status, level of education etc.
all qualitative variables are measured on a norminal scale.   For
example the qualitative variable education can be divided into six or
more groups such as no schooling, koranic, primary only, secondary,
polytechnic and university. For purpose of identification norminal
scales are assigned numerical values to indicate levels on a norminal
scale for qualitative variable as follows: 1 for no schooling, 2 for
Koranic,3 for primary only, 4 for secondary, 5 for polytechnic and 6
for university.

Ordinal Scale: An ordinal scale incorporates the features of a
norminal scale and the additional feature that observations can be
ordered or ranked from low to high.   That is, there is no notion of
numerical magnitude in the categories but are well defined, ordered,
and ranked in a way that shows that one level is higher than the other.
Exmples of norminal scales are the grading of academic staff or the
Open University.   On an ordinal scale, the academic staff can be
ranked as follows: Professor 1, Associate Professor 2, Senior
Lecturer 3, Lecturer I 4, Lecturer II 5, Assistant Lecturer 6 and
Graduate Assistant 7.

Interval Scale: An interval scale contains all features of both
norminal and ordinal scales and the additional feature that makes it
possible to specify distances between levels on the scale.   The

numerical distances between any two points on interval scale are known and can be calculated. An example of an interval scale is the average scores of students who sat for the UME in all the secondary schools in Abuja in 2002. In an interval scale it is possible to rank the average scores from lowest to highest and state exact distances measured in score units on the performance between the schools.

Ratio Scale: A ration scale contains all the features of norminal, ordinal and interval scales and in addition ratios can be formed with levels of the scale. Nearly all quantitative variables are measured on the ratio scale. Examples of ration scales are morbidity rates, death rates and population growth rates.

4.0   Conclusion

Two types of variables are distinguishable in biostatistics. They are quantitative variables whose magnitude has numerical values and qualitative variables, which are mainly categorical with no notion of numerical strength. Qualitative variables are always measured on a norminal scale, while ordinal, interval an ration scales are appropriate for quantitative variables. The four scales differ in their ability to quantify data norminal scales allow for no qualitative interpretation, while ratio scales possess the most quantitative sophistication.

5.0   Summary

In this Unit, we discuss types of variables. A distinction is made between quantitative variables and qualitative variables. The Unit also examines measurements and scales of measurement for qualitative and quantitative variables. The measurement scales are norminal scale whose levels are identified by name only, ordinal scale whose levels can be identified by name and can be ranked according to their relative magnitudes, interval scale which combines the functions of both norminal and ordinal scales, and in addition distances can be determined between levels of the scale, and ratio scales which combines the functions of the other scales and in addition ratios can be formed with levels of the scale. Norminal scale is the least advanced as a measurement instrument while ratio scale is the most advanced.

6.0   Exercise

Discuss the types of variables in social sciences

Examine the possible sources of differences in scores among a   group of individuals.

Identify and discuss scales of measurement of qualitative and quantitative variables.

7.0   Further Readings and Other Resources

Blalock, H.M. Social Statistics, 2nd ed. New Yourk: McGraw-Hill, 1972.

Selltiz, C., M. Jahoda, M. Deutsch and S.W. Cook. Research Methods in Social Relations, Revised One-Volume Edition. London: Methuen & CO. Ltd. 1959.

UNIT 3

BIOSTATISTICS: MEASURES OF LOCATION

Table of Contents

1.0 Introduction

This Unit is a follow-up to Unit 2, which focused on types of variables and scales of measurement. This Unit will discuss measurement. This Unit will discuss measures of location and examine the differences between the various measures of location.

2.0 Objectives

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

Define the measures of location

Identify the different types of measures of location

Understand the various statistical usage of the measure

3.0   Definition of Concepts

When quantitative data are collected on the characteristics of the population, summary figures are required to enable good use of the data. One of the techniques of summarizing data is the measure of location otherwise known as the measures of central tendency. Measure of location is the Arithmetic mean, the Median, the Mode, the Harmonic mean, the Geometric mean and the weighted mean. Of the measures of location, the Arithmetic mean, the Median and the Mode are the most commonly used in social and epidemiological studies. These are discussed in details below.

3.1   The Arithmetic Mean

An arithmetic mean or the average, in simple arithmetic, is the mean or average of a group observation. It is a measure of location of the group and a singular number that enables researchers or observers to assess the position in which the group is located with respect to other groups. For example if a group of 15 students sat for a competitive examination in Biostatistics and obtained the following marks 58, 34, 57, 54, 44, 57, 61, 21, 36, 57, 45, 47, 38, 48,51, the total marks scored by the group are added up and divided by the total number of the group. The figure obtained is the arithmetic mean for the group. The procedure is as follows:

58+34+57+54+44+57+61+21+36+57+45+47+38+48+51=708. When the total sum of 708 is divided by the number of students who sat for the examination, it gives a value of 47.2. Therefore the arithmetic mean or the average score for the group is 47.2

3.2 Median

The median is the middle observation or scores when the scores are arranged in the other of magnitude. This arrangement can be in the ascending or descending order. The median is the most central or the figure at the centre of the observations or scores. Let us return to scores of the 15 students who sat for the statistics examination. The scores can be arranged in ascending order as follows:
21, 34, 36, 38, 44, 45, 47, 48, 51, 54, 57, 57, 57, 58, 61 or in descending order as follows:
61, 58, 57, 57, 57, 54, 51, 48, 47,45,44,38, 36, 34, 21. The figure in the middle of the observations or score is 48. Therefore the median for the scores is 48. When two observations or scores fall in the

middle or the number of observations even, the average of the two observations in the middle is taken as the median.

3.3 Mode

The mode is the observation that occurs most frequently in a series of observations. This is describable when certain figures are frequently repeated in series of observations. From the example of the scores of the 15 students who sat for the statistics examination as shown above, the observation that occurs most frequently is 57 scored by three students. Therefore, the mode of the 15 observations or the scores is 57. It is also possible for two observations to occur most frequently, when a situation like this occurs, we have a bi-modal distribution.

4.0 Conclusion

In sample surveys or when we want to evaluate the performance of students in competitive examination, the measures of location are useful in descriptive analysis. Of all the measures of location, the arithmetic mean is the most useful as it is easy to calculate and amenable to mathematical calculation. However, certain situations arise when the other measures provide better summary measures. For example abnormal individuals among the observed may have an exaggerated effect on the mean. From the example of the scores in statistics examination, six students scored below the mean of the group, while eight other students scored above the mean by interpretation the mean is an indicator of the performance of the group. The mean is not useful in the situation where the observations are markedly skewed. Some observations may be too low or too high compared with majority of observation. In such situation the median provide a more comfortable measure of location.

The median is unaffected by abnormal observations, but it is unsuitable for work demanding mathematical calculation in the sense that if figure are elongated or reduced, the median will be altered.

The mode has a limited use because it is not easy to determine with precision when observations fall into groups. Although the mode is unaffected by abnormal individuals, the observation that occurs most frequently in series of observations may fall well below or above the group average. This can lead to erroneous conclusions about the group.

5.0 Summary

In this Unit, we discuss the various measures of location or measures of central tendency, the mean, median and mode. We also discuss the procedure

for calculating the various measures and their relative advantages and disadvantages.

6.0   Exercise

Identify and discuss the various measures of location

What are the merits and demerits of the mean, median and mode as measures of location:

Twenty students obtained the following marks in an examination: 60,69,56,75,57,65,54,51,47,65,45,38,61,44,36,48,54,25,33,37.

(a) Calculate: (i) the mean score
(ii) the median
(iii) the mode

(b) Discuss your findings

7.0   Further Readings and Other Resources

L.Ott, R.F. Larson and W. Menden-Hall. Statistics: A Tool for the Social Sciences. Third Edition, Duxbury Press, Boston, Massachusetts, 1983.

UNIT 4

BIOSTATISTICS: MEASURES OF VARIABILITY

Table of Contents

1.0   Introduction

This unit is a follow-up to Unit 3, which discussed the various measures of location, the mean, median, and mode. The three measures of location only locate the centre of a distribution of data but tell us nothing about the spread or variation of the scores. This Unit will discuss measures of variability and their use in social investigation.

2.0   Objectives

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

Define the measures of dispersion

Identify the different types of measures of variability

Understand the various statistical usage of the measures

3.0   The Range

The range is the simplest measure of data variation. It is defined as the difference between the largest and smallest scores of ungrouped data. When the data is grouped in classes, the range is defined to be the difference between the upper real limit of the highest class and lower limit of the lowest class. For example, if the scores for ten students in an examination are 45,47,48,52,54,57,64,69,72,75, the range is the difference between 75, the largest score and 45, the smallest score. In this case the range of the scores is 75 minus 45 which is equal to 30. The range can be used to determine variations in weight, height, salaries, wages, and temperature and rainfall e.t.c. The range is a good measure of variation when dealing with small sample.

3.1   Interquartile Range

The range is simple to define and calculate but not always a satisfactory measure of variability. Two observations can have the same range but could defer greatly in variation because any change in the extreme observations can alter its value. Calculating interquartile range minimizes this instability.

Therefore interquartile range is defined as the differences between the upper and lower quartiles of a set of observations. It is the set of observations that fall between the lower quartile and upper quartile of large set of data.

The range, the interquartile range and the median provide fairly good descriptions of a set of observations. Nevertheless, they do not say much about the variations in the set of observations. The variance and the standard deviation are suitable for the understanding of the variations in a set of observations.

3.2   Variance

The variance for an ungrouped data in a set of n scores is the sum of squared deviations of the scores about their mean, divided by n-1. The varioance is denoted as follows:

$$\frac{\Sigma\ (x-\overline{x})2}{n\ -1}$$

Where n is the population, and the sum of the squared deviation is

$$\Sigma\ (x-\overline{x})2$$

The variance is rarely used in general data presentation but most frequently used in technical and professional reports.

3.3 Standard Deviation

The standard deviation of a set observation is defined as the positive root of the variance. It is devised to ensure that the unit of the measurement of variability is the same as the unit of the observations. Like in the variance its unit is the square of the units of each observation. The standard deviation like the Arithmetic mean employs each of the observations in its calculation. It shows how distant each observation is to the Arithmetic mean. For example if the scores of five students in an examination are 6,5,8,7 and 4, the sum of the squared deviations of each score from the mean x is $(6-6)^2+(5-6)^2+(8-6)^2+(7-6)^2+(4-6)^2$.

Therefore the variance is $\dfrac{0+1+4+1+4}{5.1} = \dfrac{10}{4} = 2.5$

The standard deviation is $= \sqrt{\text{Variance}} = \sqrt{2.5}$

3.4 Coefficient of Relative Variation

It is often desirable to compare the variability of observations made on different items or variables, which are not measured, in the same unit. The procedure for doing this is often referred to as the coefficient of relative variation defined as the standard deviation expressed as a percentage of the mean x. The comparison can be effected by calculating the coefficient of variation in each observation and comparing their magnitudes.

The coefficient of variation can be obtained as follows:

$$\frac{\text{Standard Deviation x 100}}{\text{Arithmetic Mean}}$$

That is expressing the standard deviation of observations of each variable as a percentage of its arithmetic mean. The variable with the higher coefficient of variation now has the highest variability.

4.0 Conclusion

The mean, median and mode are three measures of location of central tendency. They are measures of location of the centre of a distribution of quantitative observations. They tell the observers little about the spread or the variation of the observations. The range, the interquartile range, the variance and the standard deviation provide measures for the description of quantitative data. The variance and the standard deviation of a set of quantitative observations provide the information that enables social researchers to compare variability between sets of observations and interpret the variability of a single set of observations.

5.0 Summary

In this Unit, we discussed measures of variability, the range, the interquartile range, the variance, the standard deviation and the coefficient of relative variation. These are measures useful in the description of qualitative sets of observations. The variance and standard deviations of a set of observations are particularly useful in comparing variations between two sets of quantitative observations and interpretation of a single set of observation.

6.0 Exercise

A group of 25 students sat for a competitive examination for admission into the HIV/AIDS Diploma Programme of the Open University and score the following marks:
58,34,57,54,44,57,61,21,36,57,45,47,38,48,51,60,61,65,51,75,58,54,51,45,25
.
Calculate
(i)   the mean, the median and the mose
(ii)   the range, the interquartile range, the variance, the standard deviation and
      the coefficient of relative variation
(iii)   write an essay on your findings.

7.0 Further Readings and Other Resources

L.Ott, R.F. Larson and W. Mendenhall. Statistics: A Tool for

the Social Sciences. Third Edition, Duxbury Press, Boston,
Massachusetts, 1983.

UNIT 5

BIOSTATISTICS: SOURCES OF STATISTICS

Table of Contents

1.0   Introduction

This Unit is a follow-up to Unit 4, which discussed the measures of
variability, the range, interquartile range, variance, standard deviation and
coefficient of relative variation. This Unit will discuss the various sources of
statistics in social and health investigation.

2.0   Objective

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

Identify the different sources of statistics

Determine the problems associated with the various sources

Understand the ways in which such data are collected

3.0　The Census

The census is the main source of population statistics in many countries. The census can be defined as the total process of collecting, compiling and publishing demographic data about every inhabitant of a particular territory. It is a sort of social photograph of certain conditions of a population at a given moment.

Periodicity and universality are important characteristics of censuses. Censuses must be taken at a well-defined period usually every ten years and must also include all members of the defined territory.

The concept of census differs from place to place. There is a *de facto* approach, whereby each individual is counted at the place where he was found at the time of a census or the *de jure* approach, whereby people are recorded according to their usual place of residence. Nigeria and Britain are examples of countries that practice *de facto* approach while the United States of America practices the *de jure* approach. In Brazil both approaches have been adopted in past censuses. It is not easy to say which is more superior to the other.　The success of each of the methods depends upon the circumstances of individual country.　In Nigeria for example, population mortality, multiple residence and homelessness of a significant number of the population made the *de jure* approach less satisfactory than the *de facto* approach. During the 1991 Nigerian Population Census, slightly more than one million for about 1 per cent of the population were reported as nomadic, homeless and transient population.

Items covered by Censuses Schedules in Nigeria 1951-1991

		1952/53	Year 1963	1973	1991
(i)	Name	X	X	X	X
(ii)	Relationship to	X	X	X	X
	head of	X	X	X	X
	Household	X	X	X	X
(iii)	Sex	X	X	X	X
(iv)	Age	X	X	X	-

26

(v)	Marital Status			X	
(vi)	Religion			X	-
(vii)	Ethnic origin /	X		X	-
	Nationality			X	X
(viii)	Home Place /	X	X	X	X
	Birth Place			X	X
(ix)	Language / Spoke				
(x)	Literacy / Education				
(xi)	Employment Status				
(xii)	Occupation				
(xiii)	Industry				

Sources: (1)   Federal Office of Statistics, Population Census of Nigeria, Lagos, 1952, 1963

(2)   National Census Board, the Report of the 1973 population Census Board of Nigeria, Lagos, 1973.

(3) National Population Commission, Census 1991 Summary, Lagos, 1994.

Coverage of censuses

The basic statistics normally obtained in a complete census are sex, age residential status, birthplace, employment, education, ethnic origin, religion and marital status. These statistics vary from place to place, from time to time and from census to census.

The United Nations recommends that the census should seek the following information:

Total population
Age, sex and marital status
Place of birth, citizenship or nationality
Mother tongue, literacy and educational qualifications
Economic status
Urban or rural domicile
Household or family structure
Fertility

How census began

Early census counts or enumeration were often related to taxation. The word census originated from the Latin censere meaning to value or tax (as in China). Another reason for census was military service. The ancient Greeks were known to count the numbers of adult males in times of war and of the general population when food was in short supply.

In the United States of America, early censuses were for the determination of political representation. Today, census information is confidential and used for statistical analysis only. Data on individuals are not published. People are less concerned that the census is connected with taxation. However, in the developed countries, many people are worried about the kind of questions asked during censuses and are concerned about maintaining their privacy.

The first modern census, that is, a continuing complete count taken accurately at regular intervals, began in Sweden in 1749. Norway and Denmark conducted general enumeration in 1769. In the United States, local censuses were conducted in Virginia in 1624, in Connecticut in 1756, and in Massachusetts in 1764. The 1787 American constitution made provision for a national census every ten years. This opened the way for the first nationwide census conducted in 1790. The census was conducted for a period of 18 months at a cost of $0.01 per capita. The census put the population of continental USA at 3.9 million.

By the turn of the 19th century, most European countries had begun to hold modern-type censuses. The first modern census took place in England in 1801, Belgium in 1829, France in 1835, Japan in 1873, India in 1881, Egypt in 1897 and in Russia 1897. Although China is rich in historical population statistics dating back to the Zhou Dynasty, nationwide censuses were conducted only in 1953, 1964, 1982 and year 2000. The 1982 census was the largest ever executed anywhere in the world. China is one of the countries that have conducted a population census in the new millennium. A total of six million enumerators were involved in the census.

In Nigeria, the first census of the colony of Lagos was taken in 1866. A rough estimate of the population by sex was made for the northern parts of the country in 1911. In the Southern parts the 1911 census was bases partly on estimates and partly on enumeration. The 1952/53 census was generally considered the first modern census conducted in Nigeria.

Subsequent head counts were undertaken in 1962, 1963, 1973 and 1991. The 1962 and 1973 censuses were rejected on account of irregularities. The 1991 population census was the less controversial census ever conducted in Nigeria. Another population census scheduled to take place in year 2001 has now been postponed till year 2004.

Between 1955 and 1964 an estimated 68 per cent of the world's population were covered by censuses. The coverage was almost universal in Europe 97 per cent, but was 62 per cent in Africa and 53 per cent in Asia. The position has since changed in Asia because China, the largest single group in that region, has joined the club of countries which have conducted a population census in the 20the and 21st centuries.

Problem that militate against accurate head counts vary from country to country and from time to time. The problems are more acute in the developing countries. Such problems are often organizational, physical, technical and attitudinal.

In the conduct of population census, the officially recognized body in charge of the census operations has encountered certain problems. Those identified are:

Organization

Engagement of functionaries hastily put together to conduct the census in their respective states without adequate training and relevant experience.
Lack of sufficient and experienced specialists to organize the conduct of the census.
Insufficient training for the different cadres of census functionaries.
Lack of uniformity in the execution of the actual head count.
Inadequate publicity and enlightenment of census operations.
Time allocated for the planning and preparation for the census not adequate for successful operation of the actual head count.
Inconsistency in the approach to the operations of head count: people who are expected to be counted in their households were counted at random at road blocks, market places, churches, mosques thus resulting into a double count and negating the concept of a **de facto** approach.
Poor census logistics.

Physical

Inaccessibility of all parts of the country throughout year round because of the difficult terrain.
Poor communication and transportation facilities.
Lack of office and storage facilities at both the states and National Census Offices.

Technical

Most enumeration areas are poorly demarcated.
Lack of adequate and up to date base maps.
Difficulties in the distribution and retrieval of completed census questionnaires.

Attitudinal

Politicization of census operations which often lead to double counting and over enumeration.
Lack of cooperation from the part of the respondents in respect of questions that are personal or conflicts with traditional beliefs and values
Lack of patriotism and sense of responsibility on the part of census enumerators who collude with members of the public to inflate census figures.

Although most of the problems reported during the 1973 were minimized during the 1991 census, allegations of double counting and under enumeration were reported, while politicization of the exercise was still a major problem. Nevertheless, the experience gained in the last two censuses is likely to lead to a significant improvement in the proposed year 2004 census exercise.

3.1   Vital Registration

While censuses describe the state of the population at a fixed point in time, vital statistics are a major source and a movie-camera record of the incidence of births, deaths, marriages, annulments, separations and adoptions. These events recorded and compiled at the time of occurrence or near their times of occurrence. In many countries of the world, particularly the developed ones, such registrations are compulsory and legal. In most of the developing countries, vital

statistics are non-existent. Where they exist, the data are not available in sufficient quantity and quality for any meaningful usage.

The registration of vital events started in the 14th century in Europe where local records or parish registers were kept by some churches. Later, the registration of vital events became state affair. However, up till today, the church authorities still have responsibility for baptisms, burials, and weddings, rather than births, deaths and marriages.

Civil and State registration systems developed in many parts of the world during the 18th and 19th centuries. Europeans who were not members of an established church were omitted from vital statistics until national civil registration was established in Europe. Civil registration was established in Norway in 1685, Sweden in 1756, France in 1792, Belgium in 1796, England in 1837 and in the United States of America in 1885.

Vital registration was first initiated in Lagos Colony in 1867 and made compulsory in 1908. A compulsory national vital registration system decree was passed in 1979 and the 1979 Federal Republic of Nigeria's Constitution charged the National Population Commission with the responsibility of establishing and maintenance of the machinery for continuous and universal registration of births and deaths throughout the country. The nationwide vital registration system is yet to take off. In limited centres where such exercises exist, the extent of coverage and the reliability of the information gathered are poor.

In many developed countries, a wide coverage of the registration of vital events has been achieved. In the developing countries, the difficulties and costs of establishing a complete registration system are so enormous that vital registration is unlikely to provide reliable demographic data in the foreseeable future.

Coverage of Information

The coverage and type of information derived from vital registration system vary greatly between countries. In some countries, over 50 different items of information may occur on the statistical report forms of births, deaths, marriages and divorces.

In the cases of births, deaths and marriages, the following minimum information are recommended by the United Nations.:

Minimum Information Recommended for Vital Registration by the United Nations

Births	Deaths	Marriages
Date and place of birth Name (if any) Sex Name and Surname of father Name and maiden surname of mother Father's occupation Signature, Description, and residence of information	Date and place of Death Name, surname Sex Age Occupation Cause of death Marital status (of deceased) Age of the surviving spouse, if any, of the Deceased.	Date of marriage Name and surname Ages Marital condition Occupations Residences at time of marriage Name, signature and description of person solemnizing the marriage

3.2   Population Registers

The population register is a government data collection system in which the demographic and socio-economic characteristics of all or a part of the population are continuously recorded. According to the United Nations. The idea of population register is expected to provide for the continuous recording of the characteristic of each individual and of information on the vital events that occur to the individual.

Population registers were first kept in ancient China, and were later adapted by the Japanese. In a number of European countries, a continuous population register is maintained to served a number of legal and administrative functions such registers may provide a continuous flow of data on vital events or a cross-section population census. It may also provide a direct information on internal movements of population.

Universal population registers, which cover the whole population, are less common than censuses or vital registration statistics. Sweden has

32

the most well established population record in the world. In all, only a few countries have registers with almost complete coverage used for demographic purposes. Out of these countries only four: Taiwan, Israel, Korea and Thailand are outside Europe. The maintenance of population registers requires a lot of resources and a reasonably accurate address system and a literate population. Keeping the register often means that everyone has to carry some kind of identification mark, and in some countries this is though to infringe on the freedom of the individual. People who move within the country are expected to register at the place of destination as well as give notice at the place of origin of the movement.

3.3   Migration Statistics

In the past, much of the information on internal migration has been obtained from comparison of successive census enumeration after allowance had been made for natural increase. Modern censuses now contain information on change of residence and place of birth, which facilitates migration analysis. It is now possible to obtain information on volume, frequency, direction and characteristics of internal migrations.

Internal migration statistics can also be obtained through sample surveys where information can be sought about the characteristics of migrants and the reasons for movement. In general internal migration data remain among the accurate of all demographic data.

International migration statistics are derived from the records of arrivals and departures at the international boundaries. Data are drawn from a variety of sources: frontiers control, port statistics, passport statistics of certain categories of travellers, local population registers, work permits for aliens. Persons crossing international boundaries usually have to produce their passport, and to complete various forms on arrival and departure.

Statistics of international migrations are available for only a small number of countries. Each country collects only the data, which it needs for its own administrative purposes. It has always been difficult to record all international movements by all countries. Several millions of Mexicans are known to have illegally crossed the border to live in the United States of America. The problem is more acute within the West African region where most of the borders are artificial. In the fifties and early sixties, several thousands from Nigeria, Togo and Upper Volta (now Burkina-Fasso) were known to

have migrated illegally to Ghana and Ivory Coast. In the 1980's several thousands of West African citizens particularly from Ghana has also illegally migrated into Nigeria. Since the economic difficulties began in Nigeria in 1987, millions of Nigeria have migrated to other countries in the world especially to Europe and United States of America, and recently to North and South African. A significant proportion of those who emigrated from Nigeria arrived at their destinations as illegal migrants.

Their host countries often do not have accurate statistics on them, hence they do not appear in the migration statistics of both their home and host countries.

Coverage of information on international migration

The information contained in the arrival and departure cards varies from country to country. Most countries normally ask the following minimum questions:

Information Contained in Immigration and Emigration Records

Immigration Record	**Emigration Record**
Name	Name
Sex	Sex
Age	Age
Occupation	Occupation
Address at point exit	Address at point exit
Address at destination	Address at destination
Port of embarkment	Port of embarkment
Mode of transportation	Mode of transportation
Duration of stay	Duration of absence
Reasons for entry	Reasons for departure
Signature and date	Signature and date

3.4   Sample surveys

It is expensive and time consuming to interview everyone in a country, hence demographers and other social scientists often use sample surveys to acquire information about a segment of the population on which they make generalizations about the entire population.

Sample surveys are frequently used to test the accuracy of census and registration data or collect vital statistics if registration is in adequate or non-existent. In the United States, the ten-year inter-censual period are filled with several surveys on demographic issues usually conducted by the Bureau of the Census and other organizations. It is customary to ask only a few questions from the entire population during censuses, while most items were administered to a 25 per cent sample of the population.

In many developing countries of the world where census statistics are incomplete or unreliable sample surveys often provide necessary information, which can be used to estimate demographic parameters. In Nigeria, the 1965-66 Rural Demographic Sample Survey was aimed at providing information that could not be obtained through the census of 1963. Similar surveys have recently been conducted to provide necessary benchmark data for estimating demographic parameters.

The World Fertility Survey(WFS) which ran through the 1970's, has been described as the largest single social science research project ever undertaken in the world since the inception of demographic sample surveys. The project involved some 350,000 women in 40 developing and 20 developed countries throughout the world. The primary aim of the project is to assist a large number of interested countries, particularly the developing ones, in carrying out nationally representative, internationally comparable, and scientifically designed and conducted surveys of human reproductive behaviour. Findings of the investigation are now available for nearly all the countries that took part in the project.

From the mid-1980 to late 1990's Demographic and Health Surveys (DHS) were conducted in several countries in the world, mostly in the developing countries to generate data for estimating demographic and health parameters. The Nigerian component of DHS was undertaken in 1990 and 1999. Data were collected from 8,999 households and complete interviews were conducted with 8,781 women aged 15-49 years across Nigeria in 1990. The 1999 Nigeria demographic and Health Survey like that of 1990 is also a nationally representative survey of 8,199 women age 15-49 years and 3,082 men age 15-64 years. The 1999 survey was designed to provide information on levels and trends of fertility, family planning practice, maternal and child health, infant and child mortality, and maternal mortality, as well as awareness of HIV/AIDS and other sexually transmitted

diseases and female circumcision. The results of demographic and Health Surveys are now being used for planning purposes.

Problems of sample surveys

Sampling consists of selecting a component or a segment to represent the entire population or a particular section of it. This process of selection creates certain problems such as sampling errors and the difficulty of making a sample truly representative. A sample contains a small number of people and it is often impossible to make generalizations about the entire populations. If a researcher is not adequately trained in statistics and research methodology, he may find it difficult to achieve representativeness in studies based on sampling. If the sample is not properly designed, it will save money, energy and enables the researcher to pay more attention to cases that are often neglected during censuses.

Inspite of all the methodological and the unrepresentativeness of sample surveys, they provide more detailed and high quality information than a census because more time and effort can be spent over each interview. A census question may indicate the number of children each woman has details about each birth and pregnancy will require several probing questions which are only possible from sample surveys.

3.5 Other Sources of Statistics

Administrative record systems of both private and public may serve as sources of demographic data. Such records include, the various social security programmes, which contained information on old age and survivors insurance, unemployment compensation and employment services. In the United States where such records are in sufficient quantity and quality they provide valuable demographic data.

The national identification system as well as the voting register may provide excellent demographic data for a large segment of the population of a country particularly the adult population if they are carefully kept. These records are available in most of the developed countries and can be used to generate demographic information.

In most of the developing countries, social security programmes and national identification systems do not exist. The voting registers that

exist in some developing countries contain a lot of irregularities and multiple entries that render them useless for any meaningful demographic analysis.

4.0   Conclusion

Nigeria is deficient in data and the census, which is the main source of information in many parts of the world has not been conducted on a regular basis. By the normal tradition of censuses, another census was due in Nigeria in year 2001 but for certain reasons, it has now been postponed till year 2004, two years behind schedule.   Vital registration that could be used as a substitute has no national coverage.   Most of the statistics being used for health planning are derived from sample surveys that are of limited value.

5.0   Summary

In this Unit, we have discussed the various sources of statistics, how they are generated, the problems associated with the collection of data from the various sources, and their relative advantages and disadvantages.

6.0   Exercise

a.

Identify and discuss the various sources of statistics in your country

Discuss the history of censuses and vital registration in Nigeria

Enumerate the items normally covered by census and vital registration schedules

Discuss the various stages, which you will adopt in population census of your community.

UNIT 6

BIOSTATISTICS

Table of Contents

1.0   Introduction

This Unit is a follow-up to Unit 5 which focused on the various sources of statistics in social and health investigation. This unit will discuss the various uses of statistics, the problems and the limitation.

2.0   Objective

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

Identify the different uses of statistics

Determine the problems associated with the various uses

Understand the limitations in the use of the various sources

3.1 Use of Census Data

Census statistics are the basic data required for planning, administrative and research purposes. There are two ways in which population components may enter into the planning process: for the distribution of goods and services and for the supply of the required manpower to administer them. Statistics are needed in the planning for education, health, housing, and employment as well as the demand for food and other essential services.

Educational

In educational planning the knowledge of the population the school-going age and sex is essential. The size of the population of children one year old is needed for planning for those who will be subsequently admitted into the day care centres, nursery schools and primary schools at age one, two and six years respectively. Data on population age- 11 years are needed for the development of primary school and for forecasting secondary school aged population; population aged 12 – 17 years and 18 – 24 years are useful for the development of secondary school and tertiary levels of education respectively.

Health

To plan for the number of persons per doctor, nurse and other paramedical personnel, the number of persons per hospital bed, requires population data on the size, age and sex distribution of the population in the country.

Health needs of the populations tend to differ according to their age and sex. The knowledge of the population age 0 -1 will aid the government in forecasting the magnitude of immunization programme for the young members of the population, childcare aid and family allowance to mothers. The size of the population of women aged 15-45 years will enable the planners to estimate the population of women of child bearing age, as well as their health need. The old members of the population particularly those aged 60 and above need special care and attention. The size of such group will enable the government to

determine whether the population is ageing and also to plan for the type of health care, housing and general welfare of this group.

Housing

The demand for housing depends largely on the size of the population, the age and sex distribution, family size and population distribution, (rural-urban distribution) and the mobility of the population as well as household incomes. An increase in population size, household size and income may create a new demand for residential places.   Housing demand varies from place to place and from time to time. In Nigerian the demand for more residential places is generally higher in the urban areas where the rate of growth of the population appears to be higher than in the rural areas.

Labour supply, manpower and employment

The supply of labour in any given population depends upon the size of the population, its age and sex structure, labour force participation rates and the level of fertility. The size of the population between the ages of 18 and 64 years will enable planners to forecast the country's employment need for persons in the labour force. Reliable population statistics will enable the planners to make necessary predictions about future employment situations in the country.

Demand for food

The   effects of changes in the number of people and the demand for food have been observed since the inception of population study. An increase in the size of population without a corresponding increase in food supply is a major cause of food shortage in many developing countries.   Where food production has failed to keep pace with population, demand for food imports have usually risen sharply.

In many developing countries, the rate of growth of food production in the last decades has barely kept pace with the rate of growth of the population.   Consequently, this has led to severe shortages and starvation.   Population statistics are important to monitor the relationship between population growth, supply and demand for food.

3.1   Use of Vital registration statistics

Vital registration Statistics can be used as follows:

To provide additional data, independent of census on measures of fertility

To check on census enumeration, particularly at the infant and young ages where under-enumeration is most common.

The sex ratio at birth provides information that can be used in population projections.

Provides legal and documentary evidence for purposes of certification and determination of age, civil status, rights and claims.

For epidemiological studies of the incidence and prevalence of diseases and the planning of health services and programmes.

3.2 Use of Population register and migration statistics

Population register can be used for a wide range of administrative matters which include: identification of persons in control of electoral rolls, selection for military service or any other national service programme and the preparation of the tax list. Statistics on social security, health, education, family
 income, housing and taxation can also be derived from population registers.

A wide range of statistics on the size, structure, composition and movement of the population can be derived from migration records. Such information is useful in urban and country planning as well as in the study of problems of population pressure and depopulation. The pattern, volume and seasonal variation in population movement are very crucial in agricultural regions. Statistics on these aspects are useful in agricultural planning, particularly in labour intensive ventures.

3.2 Use of sample surveys

Statistics from sample surveys can be used in checking the accuracy of census returns. Post-enumeration survey conducted after main population census is one of the most important ways by which the accuracy of any census returns can be verified. Information from sample surveys can also be used in the estimation of demographic components such as fertility, mortality and migration, or the total size and the spatial distribution of the population of a country or territory.

The rate of growth generated from the 1965-66 Rural Demographic Sample Surveys conducted in Nigeria was the only national figure

available for the estimation of the population until 1984 when the result of the National sample survey of 1981/82 was published. The new rate of growth from this study was used for several years as basis for estimating the population of Nigeria, until the 1990 Demographic and Health Survey and the 1991 Population Census were undertaken.

The general lack of reliable and adequate statistics on the size structure and composition of the population of any country or territory will affect the planning and development process of the countries. This has apparently been the case in Nigeria since the inception of the various Development Plans.

4.0   Conclusion

As social scientists we use statistics in several ways. We use statistics to characterize a group or to make comparison between groups. For example how many people will vote or voted for a particular party in an election or what is the voting pattern between the rural population and urban population or between Christian and Muslims. Statistics can also be used to design surveys, experiments, for planning and above all for evaluating the sucdess or otherwise of government programmes.

5.0   Summary

In this Unit we have discussed the various uses of statistics. We discussed the use of census data, vital registration data, population register and migration records, and sample survey data in educational, health, housing, labour, manpower and employment and food planning. In planning the various sources data tend to complement each other.

5.0   Exercise

Identify and discuss the use of census and vital registration data in educational and health planning
Discuss how population and vital registration statistics can be useful for planning in your community

Compare the use of census data and sample survey data for planning

7.0   Further Readings and Other Resources

I.O. Orubuloye and Folakemi Oguntimehin. The study of Human
   Populations. Centre for Polulation and Health Research,
   Ado-Ekiti, Nigeria, 2002.

UNIT 7

BIOSTATISTICS: TYPES OF STUDIES

Table of Contents

1.0   Introduction

This Unit is a follow-up to Unit 6, which focused on the various uses statistics in social and health investigation. This Unit will discuss the types of social and health surveys and the problems associated with surveys as sources of social and health data.

2.0   Objectives

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

Identify the different types of studies

Determine the problems associated with the various sources

Understand the limitations in the use of surveys to generate primary data

3.0 Research Methods in Social and Health Investigation

Research methods in social investigation have increasingly become important in the field of Social Health Sciences. Today most Universities that offer Social and Health Science courses invariably include research methods in their curriculum. The inclusion is a reflection of the growing job opportunities in the field and consequently greater number of students plan to make their careers in social sciences.

Similarly, an interest in research methods is growing among those whose jobs interests are not primarily social sciences. Students with a career of social sciences who intend to obtain administrative position in government or business must acquire some knowledge about how social facts cab be gathered. As they move upward in their chosen profession they will be faced with the problem of evaluating reports. The marketing manager, the public opinion experts, industrial relation managers and the epidemiologists e.t.c. need the knowledge of social research for interpreting and weighing reports. The growth of social sciences as areas of specialization has been accompanied by the development of basic research techniques in social investigation. The increasing attention now being placed on research methods can therefore be seen as a sign of healthy development within social sciences.

The question one may ask is if scientific approach can truly be undertaken in social studies? Many years ago, the answer will be No. Today scientific approach can be taken in the study of social facts.

The arguments against scientific approach as being done in physical sciences are as follows:

That human behaviour changes so much from one period to the next to permit scientific, exact predications.
Human behaviour is too elusive subtle and complex to yield to the rigid categorizations and artificial instruments of science.
Human behaviours can be studied only by other human observers, and these always distort fundamentally the facts being observed, so that there can be no objective procedures for achieving the truth.

Human beings are the subject of such predictions and have the ability to deliberately upset any predictions we make.

If these propositions are true then social studies have weak scientific foundation. The primary concern of a social scientist is to predict social behaviour and this is fundamental for the continuous existence of the society. Social Scientist abstracts various factors from the behaviour of other people, and thus finds their behaviour understandable. In the same vein physicists abstract from the complex behaviour of matter. It is a fact that social scientists predict the unique, this is also true of other sciences. Both common sense and scientific experience suggest that is possible to develop methods of contracting observation, abstracting adequately and reducing variability and complexity, so that social sciences can become more scientific.

Attempts precise measurements of important social force have led to increasing theoretical and conceptual clarity. Today, a higher level of methodological sophistication and greater precision in social science research are being undertaken more than ever before.

The growing methodological sophistication and technical skill of social scientists are a direct outcome of the fact that at last they see their field as having the same foundations to any other science. Experiment studies are gradually being carried out as part of social science studies where procedures that reduce bias, increase reliability and permit causal inferences about causality are adopted or employed.

3.1   Major Type s of Studies

Each study or survey has its own specific purpose but most surveys fall into a number of broad groupings namely:

To gain familiarity with a phenomenon to achieve new insight for the purpose of formulating a more precise research problem or develop a hypothesis.
To portray accurately the characteristics of a particular individual, situation or group with or without initial hypothesis about the nature of these characteristics.
To determine the frequency with which something occurs or with which it is associated with something else (usually but not always, with a specific initial hypothesis).
To test hypothesis of causal relationship between variables.

Studies that have the first purpose listed above is generally referred to as formulative or exploratory studies (and yet others do refer to it as explanatory studies) while those having the second and third purposes are called descriptive studies. Experiments are especially suited to meeting the fourth requirements, since studies testing causal hypotheses require procedures that will not only reduce bias and increase reliability but will permit inferences about causality. It could be borne in mind that many studies concerned with testing hypotheses about causal relationships do not take the form of experiments.

In practice, it is almost impossible to maintain a watertight compartment between these different types of study. Any given research may have in its elements of two or more of the functions described above. The distinction is artificial and mainly for the purpose of discussing appropriate research designs.

3.1.1 Exploratory Studies

The goals of many exploratory studies are to formulate problems for more precise investigation or develop hypotheses. An exploratory study may also seek to discover the relationships between different social phenomena. Questions in this kind of research deal with relationships between two factors such as accessibility and use of health facilities. How does accessibility affects the use of health facilities in a community or how does the level of education affects the use of health facilities? Relationships may become complex. The effects of education on use of health facilities may be a function of income. Those with high income may have more access to health facilities than those with low income.

People with higher education may have different exposure compared with their counterparts with low education. The relative importance of these different factors can be investigated. Questions can also be differently formulated as propositions about the probable or expected nature of the relationships concerned rather than as direct questions. Hypotheses may thus be derived from the question: the higher the level of education the more accessible to health facilities. The higher the level of education the greater the access to health facilities for example, would be a simple hypotheses about certain relationships. The research then results in either the verification or rejection of the hypothesis. The relationships established allow us then to explain or predict in

particular circumstances the use of health facilities of particular group of people.

3.1.2 Descriptive Studies

The primary concern of a descriptive study is to discover the characteristics form of social phenomena. The product of descriptive research is the description of classification of social phenomena, including frequency distribution. An enormous amount of social and health research has been concerned with describing the characteristics of communities.

A researcher may study the people of a particular community. He or she may be interested in their age and sex distribution, their national or ethnic composition, the state of their physical or mental health, the amount of education they have had, the number of men and women using health facilities in the community and so on and so forth.

Research may also study community facilities and their use. For example, the housing condition, the number of people living in houses with water and electricity, the extent to which health facilities are used or the amount of crime in various neighbourhoods. It may also be one undertaken to describe the structure of community organization in the area or the major patterns of behaviour.

Descriptive research may focus on estimating the proportion of people in a given population who hold certain views or attitudes or who behave in certain way: e.g. How many people will send their sick children to government or private health facilities: How many think that free health services should be abolished and user fees introduced? How many people are employed in the private hospitals and public hospitals.

Other studies may be concerned with specific predictions. For example how many people would prefer a modern health care provider to a traditional or spiritual healer?

Yet others are concerned with discovering or testing, whether certain variable are associated. For example do people who patronize private health care providers spend a large part of their

income on health care? Do more Protestants than Catholics accept family planning by contraception?

When the two types of research are compared explanatory research is usually reckoned as more valuable than descriptive. Pure description does not, in fact, meet the requirements of either explanatory science or critical theory. Description is however, an essential component of explanatory and critical research.

Descriptive and explanatory research are differentiated on the basis of the nature of the questions posed and the results envisaged, but there are also differences in the way the research is organized. Experimental methods and quasi-experimental forms of research are all used in explanatory research. The goes for the interpretation. Explanatory research will always make use of multivariate tables as long as the tables are all appropriate. Techniques of data collection, on the other hand, are equally relevant for both types of research, and standards or validity and reliability for the methods used apply equally.

3.1.3 Experimental Studies

To test a hypothesis of a causal relationships between variable requires procedure that will not only reduce bias and increase reliability but will permit inferences about causality. Experiments are mostly suited to meeting this requirement. It should be clear that many studies concerned with testing hypothesis about causal relationship do not take the form of experiments.

For example to verify the hypothesis "that higher income leads to greater accessibility to health facilities will require an experimental procedure". Hypothesis can assert a relationship between more than two variables, for example between income level, accessibility to health facilities and health treatment satisfaction. A variable, which is postulated as a necessary or causal factor in a given relation is termed an **independent variable**, while a variable that changes in response, is termed a **dependent variable.** Hypothesis formulated specifically to guide a piece of research must be empirically testable. That is the variable involved must be open to empirical observation and measurement, using the appropriate operation.

3.2   Causal Inference from Experiment

48

Commonsense thinking about causality tends to be along the line that a single event (the cause) always leads to another single event the 'effect'. In modern science there are multiplicity of 'determining factors' which together make the occurrence of a given event probable. However, both common sense and scientific thinking are pre-occupied with discovery **necessary** and **sufficient** conditions for an event.

A necessary condition is one that **must** occur if the phenomenon of which it is a **cause** is to occur. If X is a necessary condition of Y, then Y will never occur unless condition X occurs.

A sufficient condition is one that is always followed by the phenomenon of which it is a cause. If X is a sufficient condition of Y, whenever X occurs Y will always occur.

A condition may be both necessary and sufficient for the occurrence of a phenomenon. In such a case, Y would never occur unless X occurred, and whenever X occurred Y would also occur.

Since the goal of experimental study is to establish some sort of relationship between two or more variables where one is the 'cause' and the other (others) the effect, the research selects two groups: the experimental group and the control group. To ensure that the two groups are identical on the variable being measured, they are randomly selected and matched.

The time at which the dependent variable is measured provides a basis for classifying experiments in two main groups: **after only** and **before-after**

3.2.1 After-only experiments

In studies of this type, the experimental and the control groups are measured with respect to the dependent variable Y only during or after the exposure of the experimental group to the assumed causal variable X (independent variable)

The after-only, design may be illustrated by a hypothetical case of an experimental study. In this study two samples- an experimental sample E and a control sample C drawn in such a way that they were nearly alike at time 1 in regard to all factors potentially affecting the result. The presumed causal factor X was

introduced into the experimental sample but withheld from the control sample. The independent variable Y is observed in both samples at T1 and T2; and the changes occurring in the experimental sample E are then compared with the changes occurring in the control sample C; and any difference observed is attributed to the causal factor.

This can be represented with the following diagram:

Experimental Group

	Before	After
X^1	X^2
X^1	$X^{1/2}$

Control Group
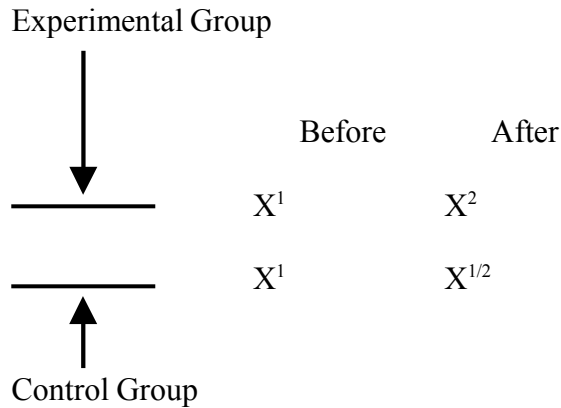
3.2.2 Before-After experiments

In addition to the measures of Y after exposure to the experimental variable, an investigator may wish to have measures of Y before such exposure.

Studies using 'before' as well as after measures of position on the dependent variable may follow various arrangements with respect to control groups.

Only one group may be used in the study, with the before measure serving as control in the sense that it is assumed to represent the level of the dependent variable in the absence of the experimental treatment.
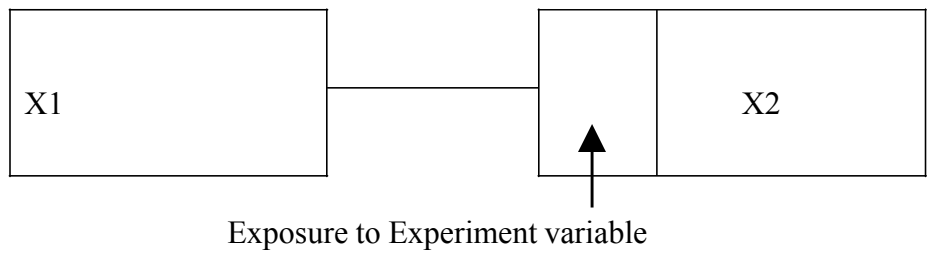
Before-after using one group

X^1	T	X^2
e		e

T = Time Interval during which the group is exposed to the experimental variable.

The before measure may be taken on one group and the after measure on a different but presumable equivalent group.

```
┌──────────────┐        ┌──────┬──────────┐
│              │        │      │          │
│  X1          │────────│      │   X2     │
│              │        │   ▲  │          │
└──────────────┘        └───┼──┴──────────┘
                            │
```

Exposure to Experiment variable

Before and after measure may be taken both on the experimental group and on one control group. In this case the two groups are measured before and after the experiment.

There may be two or more control groups against one experimental group.

M2

Experimental Group   M1   E

M = Measurement
E = Exposure

Parental
Population

Control Group   M₂   M₂
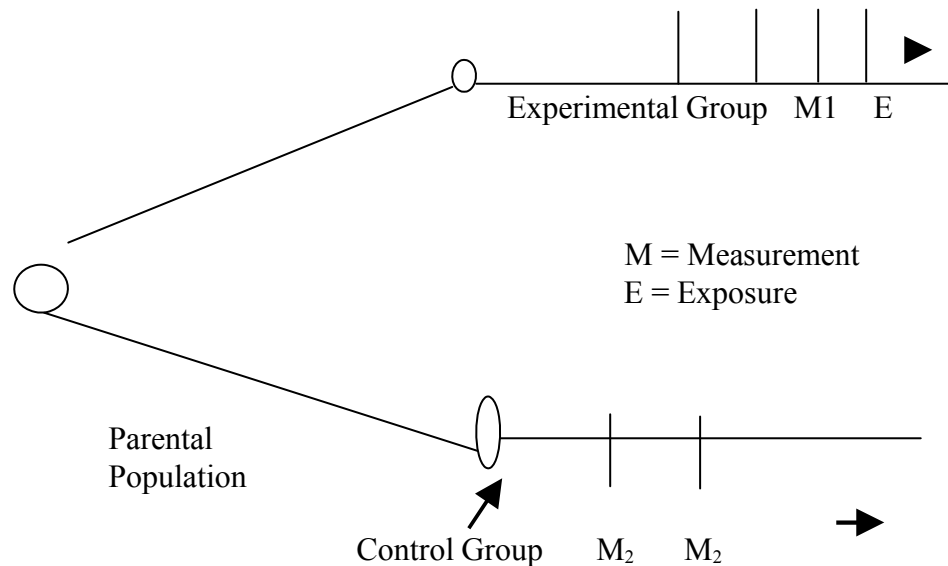
3.2.3   Problems of experimental survey:

There are several problems associated with experimental studies. These include among others:

illness/death
changes
refuse further participation
values of the researcher may affect the result
emotional factors could affect the result of the experiment
difficult to maintain strict controls on human beings under experimental conditions

4.0   Conclusion

Research methods in social investigation have increasingly become important in the field of Social and Health Sciences. Today most Universities that offer Social and Health Science courses invariably include research methods in their curriculum. The inclusion is a reflection of the growing job opportunities in the field and consequently greater number of students plan to make their careers in Social Sciences.

Attempts by social scientists to obtain precise measurements of important social force have led to increasing theoretical and conceptual clarity. Today, a higher level of methodological sophistication and greater precision in social science research are being undertaken more than ever before.

5.0   Summary

In this Unit, we have discussed types of studies. We discussed exploratory, descriptive and experimental survey. Exploratory surveys are useful in formulating problems or developing hypotheses that seek to discover the relationship between social phenomena, descriptive studies are to discover the characteristics of social phenomena while experimental studies are useful in testing hypothesis of causal relationships between variable.

6.0   Exercise

Discuss the goals of social and health research

Explain the main features of exploratory, descriptive and experimental studies

What do you understand by After only and Before-After Experiments?

7.0   Further Readings and other Resources

Selltiz, C., M. Johoda, M. Deutsch and S.W. Cook. Research Methods in Social Relations, Revised One-Volume Edition. London: Methuen Co. Ltd. 1959

UNIT 8

BIOSTATISTICS: RESEARCH PROCEDURE

Table of Contents

1.0   Introduction

This Unit is a follow-up to Unit 7, which focused on the various types of surveys in social and health investigations. This Unit will discuss the goals of research, stages of research.

2.0   Objective

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

Identify goals of research

Discuss the various stages of research

Understand the importance of the relationship between theory and

research

3.0   Goals of Research

The main goal of any research is to discover answers to questions through the application of scientific procedures. These procedures are necessary in order to increase the likelihood that the information gathered would be relevant to the question asked and will be reliable an unbiased.

Research always starts from a question or a problem of some sort. The nature of the question or problem may vary. The questions may be concerned with filling in a gap in knowledge, or with testing a hypothesis or with checking whether some proposition, which is generally believed, is in fact tenable.

In order to be answerable, research questions must have one character in common. They must be such that observation or experimentation in the natural world can provide the needed information.

Many questions of choice or decision cannot be answered on the basis of information alone, since they involve values as well as information. An example of such question is should the government provide free health care services? The answers depend on factual information such as cost of health care service, availability of health personnel as well as about individual values and initiative. This is, the question of welfare state.

Research may provide needed information; it may not provide any final answers to the questions with which it is concerned.   Despite these shortcomings, there has been a constant effort to devise procedures that will increase the probable accuracy of research answers.

For the student who is preparing for a career of carrying out research in social and health sciences, research techniques are the tools of his trade. He needs not only to develop skill in using them but also to understand the logic behind them.   Similarly, students who are preparing themselves for teaching, administration in government or business, community consultation etc require the ability to evaluate the use of research results to judge whether a study has been carried out in such a way that one can have reasonable confidence in its findings and whether the findings are applicable to the specific situation at hand.

12.1 Stages of Research

The research process consists of a number of closely related activities that overlap continuously rather than following a strictly prescribed sequence. For practical purposes and easy reference a research process consists of the following:

The initial idea

Relating the initial idea to theory

Lunching the hypothesis

Collection of data

Analysis of data

Data presentation

Feedback to theory

The details are discussed in subsequent sections

3.2   Initial Idea

This is the very first idea, which comes to the social scientists suggesting to him a new inquiry. It can happen at any time of the day or night and under any circumstance. Research topics can

trigger off from simple observations in everyday life. The incidence of malaria or STDs, HIV/AIDS in Nigeria today may be a research topic for epidemiologist.

Another starting point can be the stimulation which comes from disagreeing with what other people have said or written in a particular subject.

The government or any organization could initiate a research topic. For example, the federal government is currently planning a Behavioural Sentinel Survey on HIV/AIDS in Nigeria.

3.3 Relating the Initial Idea to a Theory

Once a social scientist has formulated the research topic, he searches the library for previous writings about the subject matter. He or she will seek classifications for basic concepts and development of a general hypothesis or ideas in more explicitly scientific fashion.

Once this task has been performed the social scientists is ready to go to the next steps, which will be one of limiting his ideas to feasible work scheme.

3.4 Limiting the Hypothesis

It may be that at this stage the social or healthy scientists have decided that the incidence of HIV/AIDS in Nigeria today is a social rather than health issue. A hypothesis as crude and general as this cannot be satisfactorily tested. To get to grips with this problem we now have to break up our general hypothesis into a number of smaller ones, which we can put to the test. For example, we may want to define what is social as well as what is health. We may suggest that the promiscuous life or poverty in Nigeria causes the prevalence of HIV/AIDS. We may even suggest that is due to sexual sins etc.

3.5 Data Collection

Collection of data in social investigation may vary widely according to the particular form taken by the research. Most social investigations take the form of field surveys. Historical and statistical source are often available to social scientists.

The method employed may be a postal questionnaire or interview schedule where the researchers conduct fact-to-face interview with the people he is studying. It could also be in the form of participant observation with or without formal interview where one fieldworker resides in a community for a considerable length of time.

A great deal of preparation needs to be made before data collection. Data should be collected because they are interesting. Good planning would reduce the amount of wasted time in the field as well as enhancement of the analysis of the data collected.

3.6   Data Analysis

The method of data collection will ultimately determine the method of analysis. A historical study may require documentary and statistical evidence to test out a number of hypotheses. In historical studies the actual quotation from records may be vital piece of information which needs to be placed in its right position to fill up gaps.

In contemporary study, particularly in social sciences, it may be necessary to analyzed hundreds of questionnaires or interview schedules. This can be done successfully and with little pain if the research worker knew in advance what he was wanting and how he intended to analyze the data before he collected it.   The researcher, who collects data on everything under the sun, usually faces a lot of problems when the analysis of all the answers has to be carried out. Questions which had no relevance to the subject in the first place are unlikely to easy to analysis at the answer stage since the researcher is seeking for answers to his hypotheses which will show him if they are to be substantiated or not.

A carefully planned research will not present difficulties at the analysis stage. Since the purpose to the answers would have been thought of in advance, and all the analysis really does is to fill in the details.

3.7   Data presentation

At this stage the initially stated hypotheses will be restated against the data which have been collected to test them and the retention or discarding of the hypotheses will take place. Results that will be 100 per cent in support of the hypothesis may be difficult to

come by. Many results, especially from demographic and health surveys may be primarily in the form of statistical tables in which percentages and averages largely tell their own stories. Written commentaries often accompany the figures and here caution must be exercised in not claiming more than the figures warrant. Sweeping generalizations must be carefully avoided. A statement must substantiated with facts.

The writer of the research report will want to draw together his results into a short statement of findings, but where some sub-hypotheses have been substantiated and others negated the problem is not easy.

The researcher will also at this stage also recount what he has left undone as well as suggest or proposed further areas of research in the field.

3.8   Feedback to Theory

The last step in the research is where the research worker makes his or her personal contribution to the discipline. No matter how small, the research states his contribution to theory and the development of the discipline in general. In doing this care has to be taken in order to avoid making statement of results that go beyond the limits justified by the data and sweeping generalization based upon nothing more than arm chair speculation.

13.0 Conclusion

The main goal of any research is to discover answers to questions through the application of scientific procedures. These procedures are necessary in order to increase the likelihood that the information gathered would be relevant to the question asked and will be reliable and unbiased.

Research always starts from a question or a problem or some sort. The nature of the question or problem may vary. The questions may be concerned with filling in a gap in knowledge or with testing a hypothesis or with checking whether some proposition, which is generally believed, is in fact tenable.

5.0   Summary

In this Unit, we have discussed the various stages of research. The procedures are interrelated starting with the research problem through data collection and analysis to report writing and feedback to theory.

**6.0	Exercise**

Discuss the goals of research

Explain the various stages of research

**7.0	Further Readings and Other Resources**

Selltiz, C., M. Johoda, M. Deutsch and S.W. Cook. Research Methods in
Social Relations, Revised One-Volume Edition. London:
Methuen & Co. Ltd. 1959

UNIT 9

BIOSTATISTICS: SAMPLING

Table of Contents

1.0   Introduction

This Unit is a follow-up to Unit 8, which discussed research procedure. This Unit will focus on the various methods of sampling and discuss the

advantages and disadvantages of each of the sampling techniques used in social investigation.

2.0   Objectives

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

Define the concept of sampling

Explain types of sampling

Discuss the use of sampling in social investigation.

3.0   Definition of sampling

If a researcher wants to collect information about some aspects of a large group, say all the University students in Nigeria, the obvious thing would appear to be a survey of all of them. But in many real life situations it is not possible to collect information about every case or the whole student population. The researcher will therefore select a part of the total population of the student. This part is generally referred to a **sample.**

A population is the aggregate of all of the cases that conform to some designated set of specifications.   We may also define a population as consisting of all people residing in Nigeria.   Similarly, we may define population as consisting of all students in the Nigeria Open University or all members of the Abuja community.

By certain specifications, one population may be included in another. Thus, the population consisting of all the Abuja community is included in the population consisting of all the members of the Nigerian community. In such cases, a population may consist of a sub-population, a population stratum or a stratum. A stratum may be defined by one or more specifications that divide a population into mutually exclusive segments.

Let us return to our example of a University community.   University community consists of students, staff and relations of staff.   The student population can further be divided into strata consisting of undergraduates and postgraduates or full-time and part-time students.   Similarly, the staff population can also be divided into academic and non-academic or senior and junior staff.   While stall relations can be divided into members of the immediate family and others.

The same University community may be sub-divided into strata consisting of males 0 – 14, 15 – 49, 50+ and females 10 – 14, 15 – 49 & 50+.

A single member of a population is referred to as a **population element.** We may want to know how certain characteristics of the elements are distributed in a population for example the age distribution, or the proportion of the element who has been to school or who belong to a particular religious organization or ethnic group.

A **census** is a count of all the elements in a population and/or a determination of the distribution of their characteristics, based on information obtained for each of the elements.

A census or complete enumeration of the University community for example will consist a count of all the students, staff and relatives of staff residing in the University community.

It is generally much more economical in time, efforts, and money to get the desired information for only some of the elements than for all of them. When we select some of the elements with the intention of finding out something about the population from which they are taken, we refer to that group of elements as a **sample.**

How closely the findings from a sample corresponds to what we would find by a comparable census of the population depends largely on the way the sample is selected.

Sampling saves time, labour and money and by reducing the numbers of cases involved. It allows for a concentration of effort on high quality information about the smaller number of cases involved. It must be recognized that statements derived from the results of sample survey are probabilistic in nature.

3.1 Probability Samples

There are two major types of sampling. They are probability and non-probability samples. Probability samples include simple random sample, stratified random sample and cluster sample. Non-probability samples include accidental sample, quota sample and purposive sample.

The essential characteristics of probability sampling is that one can specify for each elements of the population the probability that it will be included in the sample for example that each element has equal chance of being included in the sample. Probability sampling is the only approach that makes possible representative sampling plans. It enables the investigator to estimate the extent to which the findings based on his sample are likely to differ from what he would have found by studying the total population.

3.1.1 Simple Random Sample

A process that gives each element in the population an equal chance of being included in the sample is called a simple random technique. It also makes the selection of every possible combination of the desired number of cases equally likely.

For example, if a researcher wants a simple random sample of 25 cases from a population of 80 students who registered for statistics, he or she can adopt the following procedure:

Write the name of each student on a small piece of paper.

Fold each piece of paper properly.

Put all in a hat and mix them thoroughly.

Ask a blindfolded person to pick 25 pieces randomly.

By this process each person has the same chance of being selected. In a large population, the researcher can assign numbers to every body and then select a sample by using a list of random numbers. In a similar vein each member of the population has been given equal chance of being included in the sample.

3.1.2 Stratified Sample

The choice of a stratified sample means dividing the basic population into various groups (strata) and taking a separate sample from each stratum. The way the population is divided into strata is determined by the nature of the study. If, for instance, we are interested in the difference between lecturers in Universities of Technology and Universities of Agriculture the characteristics 'Technology University lecturer and Agriculture University lecturer' cab be directly selected

for stratification. Similarly, if we are interested in studying voting patterns among the Nigerian University students, the characteristic sex and age may be selected for stratification. This will yield strata such as males below 21 years of age, male 21 years and over and females below 21 years of age and female 21 years and over.

In stratified random sampling, a simple random sample is taken from each stratum, and the sub-samples are then joined to form the total sample. Stratification contributes to the efficiency of sampling if it succeeds in establishing classes that are internally comparatively homogenous with respect to the characteristics being studies – stratification ensures that all small cases are equally represented in the sample. Stratification may be employed at the stage of data analysis if this was not done initially.

3.1.3   Cluster Sample

Simple and stratification random samplings are usually expensive except when dealing with small and spatially concentrated population. The wider the area covered by a survey the more money and time will be expended on the survey. Hence large-scale survey seldom make use of simple or stratified random samples, instead they make use of the methods of cluster sampling.

In cluster sampling, a researcher arrives at the ultimate set of elements to be included in the sample by first sampling in terms of large groupings (clusters). The clusters are selected by simple or stratified random sampling methods; and, if not all the elements in these clusters are to be included in the sample, the ultimate selection from within the cluster is also carried out on a simple or stratified random sampling basis.

If, for example, we wish to take a sample of the population of the Abuja city, then we have, with the help of a map of the city, to divide the area into a number of blocks, each with approximately (but not necessarily) the same size. From the blocks we can then take the required sample, within each selected block, we can take a sample of households.

In a survey of final year students in all the secondary schools in a state, one may follow the following procedure:

Prepare a list of schools according to the number of Local Government Areas.

Classified by the size of community.

Select a simple or stratified random sample

For each of the schools in Local Government Area in the sample, list the schools and take a simple or stratified random sample of them.

If some or all the schools selected for the sample have more final year classes (or SS 3) that can be studied, one may take a sample of the classes in each of the schools.

The survey instruments may then be administered to all the children in these classes or if it is desirable and administratively feasible to do so, to a sample of the pupils.

Ideally, the procedure employed in cluster sampling moves through a series of stages hence the 'term multistage' sampling is often applied. Moving from more inclusive to less inclusive sampling units until one finally arrives at the population elements that constitute the desired sample.

3.2   Non-Probability Samples

In non-probability sampling, there is no way of estimating the chances that each element of the population will be included in the sample. In a non-probability, if there is a class of elements that have no chance of being included, by implication, there is no assurance that every element has some chance of being included. This uncertainty implies that there can be no assurance as to the precise nature of the population that is being sampled.

The major types of non-probability samples are accidental samples, quota samples and purposive samples.

3.2.1   Accidental Sample

In accidental sampling, a researcher simply reaches out and takes the cases that fall to hand, continuing the process until the sample reaches the required size. One may take the first 10th people one meets on the street and who are willing to be interviewed. A lecturer wanting to make some generalization about Nigerian University students may interview a section or all the students his or her class. Similarly, a Journalist wanting to know how the public feels on the performance of the various political parties or governments in Nigeria may interview conveniently available taxi driver, market women, teachers and others who are presumed to reflect public opinion. Although one saves, time, money and effort in accidental sampling, there is no known way of evaluating the biases introduced in such samples. In accidental sample, one can only hope that one is not being too grossly misled about the subject under investigation.

3.2.2 Quota Sample

Quota sampling, sometimes called 'representative' sampling guarantees the inclusion is the sample of diverse elements of the population and to make sure that these diverse elements are taken care of in the proportions in which they occur in the population. Essentially, the basic goal of quota sampling is the selection of a sample that is a replica of the population to which one wants to generalize. If for example, we are sampling from a population with equal numbers of males and females, and there is a sharp difference between the two sexes on the perception of the subject under investigation, and if we fail to interview males, the result of the investigation would certainly not represent the true picture of the population. In anticipation of such possible differences between subgroups, the quota sampler tries to guarantee the inclusion in his sample of enough cases from each stratum.

In practice, disproportion between the sample and the population often occurs. Some strata of the population may be over represented while some may be under represented. If this happens the inadequacy in the sample can be corrected in the analysis by employed weighting factors to the different strata in terms of their proportions in the population. This may be done by multiplying or dividing the obtained results by the appropriate figure.

Quota sampling is attractive because it is easier, quicker and cheaper than actual house to house call. Despite the advantages, there are distinct limitations, which includes:

Problems of statistical test especially the calculation of sample error, which is impossible in quota sampling because it is not based on random sampling.

Interviews may introduce bias by bending information so as to fill quotas.

People on the busiest streets are not always a good example of the general population. Taxi driver, market women are always the targets for quota sample and are usually over represented as an occupation in such sample.

3.2.3 Purposive Sample

The basic assumption behind purposive sampling is that with good judgment and appropriate strategy one can handpick the cases to be included in the sample and thus develop samples that are satisfactory in relation to one's needs. A common strategy of purposive sampling is to pick cases that are judged to be typical of the population, in which one is interested, assuming that errors of judgement in the selection will tend to counter balance each other. In practice this has not always been the case, there is no way of knowing that the "typical" cases continue to be typical. Purposive sampling is useful in forecasting national elections.

3.3 Combination of Probability and Non Probability Sampling

For all practical purposes, a researcher may combine the two approaches in one sampling design if sampling is carried in a series of stages.

The researcher may select clusters by probability sampling techniques but at the final stage select the elements as a quota sample. Thus it is possible to select a probability sample of Local Government areas in a state within each LGA a probability sampling or stratified random, sample or settlement, a quota sample of say age and sex of the elements.

An investigator may also take a probability sample of element within a non-probability sample of areas the areas are selected purposively and elements selected on the basis of a random sample.

In population and health survey, villages and town may be selected because they are typical of the settlement patterns.

3.3.1 Probability and random sampling:

There are two ways of looking at probability in so far as scientific method is concerned.

Probability refers to the likelihood that a given statement is a true statement.

The second conception of probability is essentially a mathematical one. It holds that probability expresses the frequency of the occurrence of a given event, relative to a given event, relative to the frequency of the non-occurrence of that event, in a combination that can produce either occurrence or non occurrence.

For example, if a coin is tossed into the air it has an equal chance of turning up heads or tails (since it has only two sides). If heads are designated as X and tails Y then the probability of securing heads may be stated as X = ½. The probability of securing tails is also ½. This can be expressed as follows:

X = ½N

Where N = The possible number of events. Thus the number of heads expected in 100 tosses of a coin would be ½(100), or 50 heads.

In the game of throwing the dice, the probability of throwing a 1 is equal to 1/6. Similarly the chances of throwing a 6 = 1/6. Therefore the chances of throwing either a 1 or a 6 are equal to 1/6 +1/6 or 1/3. The probability of throwing a 1, a 4 and a 6 = 1/6 + 1/6 +1/6 = ½. The higher the number of cases the higher the chances of the elements.

The advantage of such a design is that the major economies of quota sampling occur in obtaining the particular cases for the sample. It is relatively inexpensive to select the areas with which the final stage of

sampling will take place by probability sampling and one thereby gains the advantages of probability sampling. Quota samples built up in selected areas are more successful in controlling for such variables as socio-economic status (age, sex, education, income and occupation) than quota samples in which the control of these variables depends on the judgements of the interviewers. By using probability sampling to select the areas, one can gain an extra measure of security at relatively little cost.

The second example of combining probability and non-probability sample involves the opposite strategy. The researcher takes a probability sample of elements within a non-probability sample of areas. The areas are selected as a purposive sample. For example, a number of local government areas may be selected on the grounds that they are typical of their respective states either in terms of their development or in terms of their population distribution. Within each of the typical local Government areas, the researcher selects a probability sample of eligible population elements. One advantage of this approach is that if a probability sample of the population is taken, the mathematical theory of probability sampling is completely applicable and one can state the probable limits of error in the relation of the sample results to the true population. One can also generalize the inferences regarding this restricted population to the national population on the assumption that the typical local government areas are still typical of their respective states. As long as the assumption is valid, the procedure is likely to produce the most dependable sampling result with minimum cost.

The sampling plan has frequently been used in population surveys where villages and towns are selected because they are typical of the settlement patterns or because they are typical of certain basic criteria set out for the investigation. The population element within the villages selected by adopting a probability sampling approach.

3.3.2 Application of Non-Probability Sampling

As indicated in our previous discussion, the major advantages of non-probability sampling are the convenience and economy. One of the disadvantages is that it does not provide any basis for estimating how far the sample results are likely to deviate from the true population figures. Despite this shortcoming, researchers have largely relied on the use of this approach. The argument being that the superiority of probability sampling in non-probability sampling exists only on paper and that there is a difference between the sampling plan and its actual

execution. For example, Interviews may fail to follow their instructions in selecting respondents, or they may omit some of the questions in interviewing some of the respondents.

Some of the selected cases may refuse to be interviewed or not being available for interview compromises may be made by allowing interviewers to substitute other respondents when those designated for the sample are not found at home. The sample actually obtained may, hence not be the probability sample it was planned to be. There are circumstances where probability sampling is unnecessary or inappropriate. These include:

During exploratory survey when the researcher wants to gain insights into a particular phenomenon. In such a situation, one selects a purposive sample.

The second example is a situation when a researcher samples ideas rather than for the estimation of population values – the field of market research known as motivation research falls into this group. The primary aim of motivation research is to find out something about motives, attitudes and associations. The results of such studies are used by advertising agencies which make use of them in developing advertising campaigns- Accidental samples or with purposive samples selected in such a way as to maximize the likelihood of differences among the element in the sample. The primary concern is to collect the relevant ideas to transmit to the advertising agencies, not for correct estimates of population distributions.

Sometimes there is no alternative to non-probability sampling. If one is trying to find out something about a special group or a special case, for example about the attitudes of people in the other side of the **"Iron Curtain"** one has no realistic choice but to rely on informants who have recently been there. Similarly, when a researcher wants to reconstruct a picture of a dying culture one has no choice but to rely in relatively articulate informants for certain types of information.

Another special case justifying the use of non-probability samples is when it is necessary to balance one consideration against another – for example, a better sampling design against a more sensitive method of data collection. The study of the characteristics of patients in psychoanalytic treatment where analysts serve as informants requires a high degree of anonymity

hence accidental sample becomes a necessity both for the patients and the informants.

Many studies in behaviour science are carried out on accidental samples of subjects. The data are later treated in a manner that is appropriate only to probability samples. Statistical tests of significance, which presuppose random sampling are applied to the data.

3.4   Method of Sample Construction

During our discussion on random samples, we emphasized that a random survey will never be completely representative; it cannot go beyond representation of the basic population approximately. If for example, the proportions of the sexes in the student population of the Nigerian Open University is 40:60, then in a random survey, we may find a proportion of 35:65. In this situation, it is important to know something about the size of the sampling error OR it is important to be able to say with a certain degree of confidence (perhaps 95 per cent), the true value lies in an interval between say 5 per cent under and 5 per over the value found in the partial survey.

The central requirement of statistical sampling theory is that each unit of the basic population must have a **calculable** (in the case of the simple random sample, an equal) chance of being included in the partial survey. In doing so the following rules must be observed:

The basic population must be physically or symbolically present and capable of manipulation (i.e. individually shifted around and arranged at will). This requirement is, incidentally very difficult to fulfill in social investigation. For example, the population of Nigeria is not either physical or symbolically ever completely assembled, so that a researcher can manipulate it for the purpose of taking a sample. The inhabitants of the City of Abuja or the students at the Nigerian Open University too are never to be found completely assembled. Notwithstanding, in a population census or a register of students, they are however symbolically present are therefore can be manipulated.

The units of the basic population must be well mixed. If, for example, an index of students is drawn up according to year of study or course of study, and we take the first 500 cards as a sample, then we are choosing only students in their first or second year or students in one or two faculties. In doing this, the principle of equal chance being selected is violated. The index must be well mixed. Although

this impossible in practice, it can be done symbolically by using a table of random numbers.

Other simple methods generally known as systematic random sampling, can also be applied, say for example choosing every tenth card or selecting every person whose name begins with a certain letter O or A or selecting all people who were born on August 20th or March 15 or choosing by a final digit in the case of cards number serially, choosing all those ending with a certain digit say 4 or choose the last two.

Each unit must be included in the index representing the basic population only once (or all units must be included with the same frequency). For example, if we want to draw a sample of parents with children in a public school, parents with more than one child is likely to be over represented since they have a greater chance of appearing in the sample than parents with only one child. To observe the principle of each unit being included in the index representing the basic population only once the card of the second, third or more children must be excluded. Only then can the register of school children be a good symbolic representation of the population of parents and thus useful as a sample base.

4.0   Conclusion

Sampling provides the mechanism by which we can obtain information for a large and diverse population. It is generally more economical in time, effort and money to get the desired information for only some elements than for all of them. If appropriate sampling plan is adopted the results can be generalized to the entire population, and any decisions based on the findings are likely to have the desired impact on the total population.

5.0   Summary

In this unit, we discussed the various forms of sampling, probability and non-probability sampling techniques, and their relative advantages and disadvantages. We also discussed the conditions under which the sampling strategies can be combined and special applications of non-probability sampling in social research.

6.0   Exercise

Define the concepts of sampling and population element

Identify and discuss the various types of sampling

Compare and contrast probability and non-probability sampling techniques

Discuss the sampling techniques that you will adopt in the study of use of health facilities in your community

7.0   Further Readings and other resources

Selltiz, C., M. Jahoda, M. Deutsch and S. W. Cook. Research Methods in Social relations, Revised One-Volume Edition. London: Methuen & Co. Ltd.

UNIT 10

BIOSTATISTICS: DATA COLLECTION

Table of Contents

1.0   Introduction

This Unit is a follow-up to unit 9, which discussed sampling and the various methods of sampling procedure. It will discuss methods of data collection the various sources and the advantages of the various methods.

2.0   Objectives

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

Identify the various methods of data collection.

Explain the similarities and differences between the various methods.

Discuss the use of the various methods.

3.0   Documentary sources

There are three main methods by which data can be collected. These include: (i) documents (ii) observation and (iii) the questionnaire or the interview schedule. It is customary in research to distinguish between the sources of the documents by classifying them into primary and secondary.

Primary sources provide data gathered at first hand; that is to say they are original sets of data produced by the people who collected them. They differ from secondary sources, which are data got on a second-hand;-set of data not collected at first hand but, culled from other people's original data. The distinction between primary and secondary source can be made even more useful if we adopt a further division of documents between what John Medge after Gottschalk, calls records and reports.

Primary   **Secondary**

Compiled at the time by the Writer: e.g. Court records, Census of population	Transcribed from primary contemporary sources: e.g. research report bases on assistants fieldwork
Newspaper report	Historical study using actual documents
Contract	Statistical research based on census data.
Letters	Research using, other peoples' correspondences
Tape-recording Currently in writing by self	Written on the spot by somebody else.

Retrospective:

Primary   Secondary

Compiled after the event by	Transcribed by primary

the writer: e.g. Personal diary Autobiography	retrospective sources: e.g. Research using diaries or auto-biography written afterwards by somebody else.
Report on a visit to a given institution	
Written afterwards by self	

The distinction between records and reports is that record is primarily concerned with an event that is currently taking place e.g. an act of Parliament, a contract between two people, the annual statistics for a government department, taking of the record of vital events, (births, deaths, and marriages)

3.0.1   Official Records

Official records are supposed to be the most reliable sources so long as we can trust the writer.   The verbatim parliamentary record of what is said in the Senate and the House of Representative is possibly the most trusted document one can find.   The proceedings of Judicial cases would also occupy such an important position.

3.0.2   Newspapers

Newspaper reports, where a reporter was present at the scene, might be thought to be valuable, but unfortunately it has been shown only too often how reliance can be placed on them.   Many reporters do not have shorthand and use their own personal notation system.   Their reporting of speeches is likely to be inaccurate in many ways, sometimes embarrassingly so.   Also a reporter can only extract a tiny part from any total occurrence, and in many cases he extracts what will be eye-catching and provocative therefore could not be a good source of data.

3.0.3   Official Statistics

These include

The census of population

The monthly and annual digests of statistics

The monthly statistics of HIV screening

The annual statistical reports of the various ministries and other

national bodies e.g. the education, health and prison departments

The unfortunate aspect of our data situation in Nigeria is that people do not trust official government statistics neither do people trust the use made of them by politicians and others.

3.0.4 Other Sources of Documentary Sources

Diaries	-	recorded at the time of event
Memories	-	written by self from recollections
Autobiographies	-	written by self
Biographies	-	written about by self or by others
Correspondence	-	private and public

Historical document   - historical document refer particularly to events of the past about which the main (and probably only sources of information is documentary, the participants now being dead – current historical document do include people who are still very much alive e.g. History of the Nigerian Civil War.

Story of Nigerian Independence

Most work in community studies begins with a historical development of the settlement.

The studies of organizational structure such as the health care delivery system or the structure of a modern hospital

3.1   Observation

A wide variety of data collection procedures are included under the heading of observation. They are participant observation used principally in anthology, social and cultural anthropology in studying relatively small-scale, primitive socio-cultural system and Non-participant observation- observation of group behaviour under controlled conditions.

All types of observations are concerned with concrete behaviour, conduct and symbolic interaction in straightforward or more complex social situations.

The type of method of observation employed on any occasion depends on the nature of the research topic. Three types of problems can be distinguished.

It may be a matter of producing a straightforward description of behaviour patterns and forms of social interaction occurring in a particular socio-cultural system, without following up any explicit theoretical interest. The research may be directed explicitly toward analyzing the empirical relationships between certain sequences of behaviour and situational conditions.

It may be a question not only of describing a certain behaviour pattern in a particular socio-cultural system, but also of explaining it in terms of its conditions and its objective consequences in the light of general theoretical statements.

3.1.1 Unstructured Observation/Participant Observation

This approach is used mainly in social anthropology. In this form of observation, the observer takes on, to some extent at least, the role of a member of the group and participates in its functioning.

The basic question the observer must face is what should be observed? In order to narrow down and to facilitate observation what an individual could cope with. The following checklist suggests directions of observation they may otherwise be overlooked:
The participants – who are the participants how are they related to one another, and how many are they – the following will help to solve the problems of identification: age, sex, official position (teacher, chief, senator, honourable and spectator etc). In terms of relationship one would like to know whether they are strangers to one another or they are members of the same community.

The setting- A social situation may occur in different settings – e.g. in church, schools,- clubs houses etc. What kind of influence has the social setting on the behaviour of the participants.

The purpose – is there some official purpose that brought the participants together? Are they together by accidents – Do they all have the same goals? Are these goals contradictory?

The social behaviour – here an observer is faced with reality of the situation – What do the participants do, how do they do it- What was the stimulus or event that initiated it – What external or internal motivation do they have. Frequency and duration – when did the situation occur how long did it last. Is it a recurring type of situation or unique? If it occurs, how frequently does it occur? What are the situations that give rise to it? How typical of such situation of the one being observed?   Is such a situation likely to give rise to such an event?

The list is not meant to apply in totality to every situation observed. In most cases it is impossible to obtain enough clues to permit such a detailed description.   The events may be too rapid to permit evaluation of all aspects.

3.12   Recording of Events

The most appropriate thing to do is for the observer to record all the events on the spot and during the event. This result in a minimum of selective bias and distortion due to memory failure.

This may not be feasible in all cases – because it may disturb that naturalness of the situation or arouse the suspicion of the persons observed.

Constant note taking may interfere with the quality of observation. The observer may easily lose relevant aspects of the situation if he divides his attention between observing and writing.

The main problem with participant observation is the problem of objectivity. An observer is likely to develop friendly relations with some of the people he is studying. He may find himself personally concerned with a story told by an informant. Full recording helps to restore objectivity. This can now be achieved through the use of tape recorder.

3.13   Structured Observation/Non-ParticipObservation

Structured observation focuses on specific aspect of behaviour, which may take place either in field setting or in controlled experiments within a laboratory setting.

Used mainly in studies designed to provide systematic description or to test causal hypotheses – For example structured observation will be suitable in studying management/worker relationship, students/teacher relationship.

There is no one best method of recording observations- The most frequently used system of recording is one that provides the observer with a number of duplicated sheets containing the list of categories to be coded and the cells in which they are to be marked.   Sound recordings and motion pictures have been used to record the over-all nature of an event or to code certain actions of members or of the group in terms of the frame of reference provided by the entire want.

Non-participant observation cannot be carried out in natural situation without the knowledge of the people under observation.   The major advantage is that the observer can devote his whole attention to observation and make notes during its progress.

3.2   Interview and Questionnaire Approach

While observation methods are primarily directed toward describing and understanding behaviours as they occur, the interview and questionnaire approach take the form of verbal report of information about the experiences of the individual.   This approach is used to collect individual data, which are collated as group data in the course of interpretation.

Although it is the individuals that are interviewed, the results are collated according to classes, groups or types of individuals, which are of interest to social scientists including epidemiologists and demographers.

By using interview approach, we are able to ascertain both subjective and objective facts.   Distinction can be made between questions about fact and questions about opinion.   Questions about facts include verifiable facts, which may refer either to the subject himself (e.g. his age, sex, marital status, family size, etc).

Question about opinion aimed to discover subjective attitudes.   These include questions about opinions and value judgements concerning

objective state of affairs to attitudes, desire, feelings, motives and norms of individual behaviour.

The interview method is probably the most frequently used research method in epidemiology and demography, and for ascertaining subjective situations, which can be expressed verbally as the most appropriate.

Although both interviews and questionnaires place heavy reliance upon the validity of verbal reports, there are important differences between the two methods.

In an interview, since the interviewer and the interviewee or respondents are both present as the questions are asked and answered, there is opportunity for greater flexibility in eliciting information. In addition, the interviewer has the opportunity to observe both the subject and the total situation to which he is responding. In a questionnaire approach, the information one obtains is limited to written responses of subjects to pre-arranged questions.

3.2.1   Advantages of Interview

In Nigeria the level of illiteracy is very high probably about 70% compared with just about 10% in the USA. Thus interview method becomes the most appropriate means by which the population can be reached;
Interviews can be used with all segments of the population

Yield a much better sample of the general population Yield

very high returns compared to mail questionnaire Interview

is more flexible – there is opportunity for
repeating or rephrasing questions to make sure that they are understood.

Interview offers better opportunity than questionnaire to appraise validity of reports. The interviewer is in a position to observe not only what the respondent says but also how he says it.

Interview is more appropriate for revealing information about complex, emotionally laden subjects or for probing the sentiments that may underline an expressed opinion.

83

3.3   **Advantages of Questionnaire**

The advantages of questionnaire are:

Questionnaire is likely to be less expensive procedure than the interview.

It requires much less skill to administer than an interview.

It can be sent through the mail whereas interviewers cannot.

It is an impersonal means of obtaining personal information.

It standardized wording, order to questions and instructions for recording responses. It ensures some uniformity from one measurement situation to another.

It enables the respondents have greater confidence in their anonymity, and thus feel freer to express views they fear might be disapproved or might get them in trouble.

The major disadvantages of questionnaire are low response rate and high number of no response or don't know answers.

3.2.3 Factors that may influence the Rate of Return of Mailed Questionnaire

The factors that may influence the rate of returned questionnaire are:

The sponsorship of the questionnaire

The attractiveness of the format of the questionnaire

The length of the document

The accompanying letter of introduction

The ease of filling out the questionnaire and mailing it back

The inducements offered to reply

The nature of the people to whom the questionnaire is sent

The postal system

3.4   Questionnaire Design

There are certain rules to guide designers of questionnaires and interview schedules. According to Bowley, these rules include among other things:

Ask for the minimum of information required

Make sure that questions can be answered

Make sure that questions will be answered truthfully

Make sure that questions will be answered an not refused

There is always a number of questions which could be asked in a survey because they seem interesting but interest is not enough.   A question to be included should be relevant to the problem being studied.

Questioning is particularly suited to obtaining information about what a person knows, believes or expects, feels or wants intends or does or has done, and about his explanations or reasons for them.   Question contents do not always fall neatly into one another (content type). The distinctions are merely artificial.

The major types of question content are as follows:

3.2.4.1 Content aimed mainly at ascertaining facts

These sets of questions include:

Persons age, education, religion, income nationality, state of origin, marital status, occupation etc.
Questions about the characteristics (behaviour beliefs, feelings, desires, intentions etc.) of persons who are known to the respondents, such as family, friends and colleagues.
Reported facts must always be evaluated in terms of credibility.

3.2.4.2 Questions about beliefs about what the facts are

The investigator decides or may wish to learn what people believe to be the facts. For example asking a respondent to indicate whether certain statements are true or false.

In this respect the respondents answer is not used to establish what is objectively true but rather to provide a picture of his beliefs. This is of course on the assumption that the respondents have any beliefs or information relevant to the topic under investigation. The distinction between an inquiry into **facts** and an inquiry into **beliefs** must be kept in mind by the investigator.

3.2.4.3 Content aimed mainly at ascertaining feelings

A person's beliefs about what the facts are will often give very clear indication of his feelings and his desires. The most common method of investigating feelings in surveys is to include items that bear directly in various possible emotional reactions. These may include fear, distrust, disgust, contempt, hate, envy, sympathy, and admiration. When a person is sick the ultimate wish is to get well quickly. When there is a football match between Nigeria and France most Nigerians will want the Nigerian team to win irrespective the quality of the team and state of preparedness.

These desires may be based purely on emotional feelings. An investigation of emotional reactions, if so to provide a full picture, must uncover not only the individual's feelings but also the circumstances in which the feelings are likely to be aroused. For example, instead of asking how do you feel about the health care facilities in your community, a researcher may ask have you ever use the health care facilities before? If yes why and what are the motivations.

Other questionnaire contents may include, content aimed mainly at discovering standards of action, present and past behaviour, feelings and policies.

3.3   Type of interviews and Questionnaires

The form of interviews and questionnaires may vary widely from survey to survey. Interviews may range from the rigidly standardized; both the questions and the alternative responses   permitted   the subject are predetermined, to the completely unstructured ; in which

neither the questions to be asked nor the responses permitted the subject are determined before the interview.

3.3.1   Standardized Interviews and Questionnaires

In order to ensure that all respondents are replying to the same questions, the format of the question must be the same. For example, the question would you like another census in year 2003? Is not the same as do you think another census in year 2003 would be desirable? The answers to these two questions may not be comparable. Differences in question order can also influence the meaning and implications of the given question.

3.3   Fixed alternative Questions

Fixed alternative questions or closed or pre-coded questions are ones in which the responses of the subject are limited to stated alternatives. These alternatives may be simply Yes or No or they may provide for indicating various degrees of approval or agreement or they may consist of a series of replies of which the respondent picks one as being closest to his position. For example, what is your level of education? The answer may be pre-coded as follows:

>   No schooling
>   Primary
>   Secondary
>   Teaching training
>   University

>   Or Monthly Family Income
>   Below N10,000
>   N10,000 – N20,000
>   N21,000 – N30,000
>   N31,000 – N40,000
>   N41,000 – N50,000
>   N51,000 – N99,000
>   N100,000 and above

In another vein a question cold be asked as follows: As you probably know, the Nigeria senate is currently debating abortion bill? How do you feel about this?

>   Strongly approve

Mildly approve
Undecided
Mildly disapprove
Strongly disapprove

Questions of this type are basically the same whether they are used in interviews or in questionnaires.

3.4   Open Ended Questions

Questions of this nature are designed to permit free response from the subject rather than one limited to stated alternatives. The respondent is given the opportunity to answer in his own terms and his own frame of reference.   Examples of open-ended interview questions are as follows:

Now that we have been practicing presidential system of government in Nigeria for three years, I wonder if you will tell me how you fell about it.

What do you like most about the new system?
What do you dislike most about it?
What are the advantages over the parliamentary system?
What are the disadvantages?

3.5   Advantages of Closed or pre-Coded Questions

These are:

When used in questionnaires, both the questions and the order in which they are presented are predetermine. It is impossible to ask any supplementary questions.   When open-ended questions are used in standardized   interviews,   the   questions   and   their   order   are predetermined, but interviewer can repeat the questions if the reply is not to the point or ask probing question to clarify issues.

Standardizable
Simple to administer
Quick and relatively inexpensive to analyze

Disadvantages:

May force a statement of an opinion on an issue which the respondent does not have an opinion.

No choice of positions on an issue because they do not allow for qualification.

Omission of possible alternative responses may lead to bias.

May conceal the fact that different respondents make different interpretations, some of which may be quite different, from those intended by the interviewer.

3.3.5   Open-Ended Questions

The disadvantages are:

Difficult and expensive to manage, the categories for analysis must be built up, coders must be trained, and the responses must be coded into one of the categories before they can be tabulated and statistically analyzed. The analysis of open-ended questions is complex and often troublesome.

Advantages:

Open-ended questions are useful to examine complex issue, when the relevant dimensions are not know, or when the interest of the research lies in the exploration of a process or of the individuals formulations of an issue.

3.4   How to ask Questions

Asking questions to get valid answers is a skilled and sensitive job requiring knowledge of the environment in which the questionnaire is to be filled in or the interview conducted. It requires knowledge of the likely impact of the questionnaire on the respondent.   It requires sensitivity to the symbolic sophistication of humans, non-verbal as well as verbal.

Asking too many questions is a bad starting point for any questionnaire or interview schedule.

Do not ask respondents for information, which is readily available from their sources.

Ask only relevant questions e.g. it would be disastrous to the success of an interview for the interviewer to ask for instance the sex or colour of a respondent in an interview.
Where observation is sufficient it is silly to verbalize a question.

Avoid certain questions in some other situations e.g. a survey of University students may well be done on a basis of a stratified sample in which the year of study, respondents are already from actual records. In such a situation do not ask the information you can extract from the record from the respondents.

Avoid complicated questions- particularly those that require detailed probing into ones past history.

Do not over-estimate the intelligence of your respondents when asking questions.

Avoid ambiguous questions- questions that may result in the respondent being able to make more than one interpretation of the questions.

Leading questions- These are questions which, in the very way they are put, tend to influence the answer that a person give.

Double questions- Questions like, how do you go to cinema or theatre are too two questions that require two answers.

Jargon and Technical terms.

Emotional questions- Questions asked must take care in such delicate areas as sexual behaviour, religion, politics and social class.

3.5 Other Points Relevant to Good Administration of Interview and Report

Whenever questions are to be asked and a choice made from a limited list of answers it is a safeguard if they are out in advance. Pilot studies are essential for ensuring that the responses offered as possible answers actually to exhaust all the possibilities. Some questions may be found useless as the range of answers will be limitless.

Pilot study could be in the form of possible questions with free-ranging unstructured questions or open ended inquiry may form a first pilot study and the actual testing of questions a second stage without any pilot stage, the actual research is likely to address unsuitable questions to bewildered people.

A skilled interviewer may be able to sustain interest and cooperation through a long session. Although sessions of six to eight hours have been achieved lengthy questions have not achieved very much.

The span of attention of children increases only slowly with age. Old people get tired easily. Head teachers, businessmen and the upper middle class are impatient with any for that may be time wasting. Besides, there is the general problem of apathy.

Questions can be too technical or complex. Concept that is of day today use will achieve more results than technical ones. In long questions asking for a choice between alternatives, the last is more often chosen because the first has been forgotten.

When questions are asked about life in schools, old peoples' homes, prisons or other relatively closed organizations, they may be influenced to give answers within a particular context. This may result in right but not real answers.

4.0   Conclusion

Data collection is an important step in research. The method of data collection depends on the nature of research being undertaken. There are three main sources of data collection: documentary, observations and interviews. Documentary sources are useful in the study of the history of organizations, while observations are used in the study of small-scale societies and interview for large scale-societies. The type of questions and the ways questions are asked depends on the type of investigation being carried out.

5.0   Summary

In this unit we discussed the various methods of data collection. We also discussed the ways in which they can be used, the advantages and disadvantages of each method. We also discussed types of questionnaire, interview schedule and how to ask questions.

6.0   Exercise

Identify the main methods of data collection

List the types of question you will ask in a survey of use of health facilities in your community

Explain the differences between open-ended questions and fixed alternative questions.

7.0   Further Readings and Other Resources

Selltiz, C., M. Johoda, M. Deutsch and S.W. Cook.
    Research Methods in Social Relations, Revised One-Volume Edition.
    London: Methuen & Co. Ltd. 1959

UNIT 11

BIOSTATISTICS: DATA ANALYSIS

Table of Contents

1.0   Introduction

This Unit is a follow-up to Unit 10 that dealt with data collection. After data have been collected the data is analyzed in the form in which the users can understand them.   This Unit therefore describes the procedure for data analysis with emphasis with emphasis on establishment of categories, categorization and tabulation of data.

2.0   Objectives

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

Understand the procedure for establishment of categories.

Explain categorization of data

Discuss tabulation of data

Explain statistical analysis of data

3.0   The Establishment of categories

Categories are established to summarize the varied responses obtained in a research work.   Questions often asked can either be open or closed ended. Open-ended questions are such that respondents are allowed to respond freely to issues under investigation using their languages of expression.   Close-ended questions are questions in which respondents are provided fixed-alternate responses in which respondents are asked to choose answers that are close to their positions.

Let us return to our earlier example on the presidential System of Government. In a survey of attitude to presidential System of Government in Nigeria, a sample of the population was asked the following questions: "What do you think about the Presidential System of Government?"   The following answers were obtained from the respondents:

> I like the Presidential System.
> I do not like the Presidential System.
> It is an expensive system of government.
> The president is too powerful.
> There are separation of power between the three arms of government.
> The State are too independent of the Centre.
> I wish the nation could return to the Parliamentary System of Government.

The answers can be more varied than the above. For the responses to be used in answering the research questions they must be grouped into categories. In order to decide on the relevant categories, the established principals of classification must be followed. These principles are:

The set of categories should be derived from a single classificatory principle.

The set of categories should be exhaustive, that is, it should be possible to place every response in one of the categories of the set.

The categories within the set should be mutually exclusive, it should not be possible to place a given response in more than one category within the set.

In the application of the above principle to the answers to the question on presidential System of government in Nigeria, two major categories can immediately be identified:

Those who like the Presidential System of Government.

Those who do not like the Presidential System of Government.

A set may consist of more than two categories, provided the rules are not violated. Each of the answers may stand on its own as a category, while two or three answers can be grouped together to form a set.

In some situations the establishment of categories is considerably more cumbersome than in the presidential System of Government example, especially if such studies do not start with explicit hypothesis. In this example the question is simple and the answers are also straightforward hence the establishment of categories are less cumbersome.

3.1 The Categorization of Data

Coding is the technical procedure by which data are categorized. In this way, the raw data are transformed into symbols that may be tabulated and counted. The symbols are usually numerals. Coding or categorization of data is often done after the data has been collected. The persons who do this are referred to as coders. The questionnaire or survey instrument may be done in such a way that allows the respondents to assign code to their responses during interviews. This is possible in questions in which fixed alternative answers have been supplied by the researcher.

The application of fixed alternative, otherwise know as closed ended or pre-coded answers has been discussed in Unit 10.

The reliability of coding is affected by the competence of the coders. Therefore training of coders usually precedes the coding exercise and periodic consistency checks are necessary to ensure that coders do not become careless as they become more experience.

3.2 Tabulation

Tabulation is an essential part of the technical process in the statistical analysis of data. The major operation in tabulation is counting to determine the number of cases that fall into the various categories. Tabulation may be done entirely by hand if the quantity of the data is small and the number of category sets to be counted is small. When the quantity of the data is large and the number of category sets is also

large tabulation is often done by machine. Both hand and machine tabulation pre-supposes that the data have been coded and the coding has been thoroughly checked. The invention of computers has made the tabulation of data less cumbersome and error free than has been the case several decades ago.

Data collected for a research purpose may be voluminous to the extent that they are not easily comprehended. Data are often collected from a representative sample of respondents that may run into their several thousands. After gathering such data, the researcher's task is to seek a way of reducing and simplifying the cases into detailed form in which important characteristics of the respondents are brought out. However, reducing and simplifying data are done in such a way that the original meaning and interpretation of the data are not altered.

The term marginal is commonly used to refer to simple counts of the frequencies with which the various categories in each set occur in the data. The terms cross-tabulation or breakdown are often employed to refer to the tabulation of the number of cases that occur jointly in two or more categories.

3.3   Statistical Analysis

Data collection is a pre-eminent activity aimed at obtaining necessary information to satisfy some decisions objectives. This broad view of data collection tends to include the varying activities of most government agencies that are charged with the responsibilities of collecting data on regular basis albeit in the anticipation that demand will be made by a user at a later date. Those involve in data collections have much to do with statistical activities involved in the process of constructing social indicators.

In a sense, statistics was used to designate collection of data in relation to matters of special interest to a nation-state such as birth, deaths, population counts, tax returns and so on. When data are obtained in voluminous character, statistical techniques are used to translate them into a more understandable piece. The term statistics have three notable connotations.

Numerical information – summary of numerical information
A discipline- the scientific method of making decision under uncertainty when numerical data and calculated risks are involved.

Branch of mathematics- statistics is seen here as a branch of mathematics that deals with quantitative data. It is the science and art of obtaining or analyzing such data in order to make sound inferences in the face of uncertainty.

In analyzing data, statistics can either be descriptive or inferential. Descriptive statistics relates to the measures such as tables, charts, graphs, histogram, charts, frequency polygon, mean, median and mode.

Descriptive statistics present a snap short of an event. The application of description statistics has been discussed in Unit 3.

Measures of variability, the range, interquartile range, variance, standard deviation, and co-efficient of relative variation provide useful measures for the description of quantitative data. They enable social researchers to compare variability between sets of observations and interpret the variability of a single set of observations. A detailed discussion of measures of variability of a single set of observations. A detailed discussion of measures of variability is contained in Unit 4.

4.0   Conclusion

A social researcher turns his or her full attention to data analysis after data have been collected. Data analysis consists of a number of closely related operations whose purpose is to summarize the completed observations in a manner that they provide answers to the research questions. The procedures used in data analysis covers the establishment of categories, the application of categories to the raw data through the process called coding, the tabulation of individual responses and statistical analysis of the data.

5.0   Summary

This Unit discusses data analysis with emphasis on the establishment of categories, which have their own distinct characteristics. It also discussed coding the technical procedure by which data are categorized. Through coding the raw data are transformed into symbols that can be calculated and counted. We also discussed tabulations, that is the counting to determine the number of cases that fall into the various categories. The last section in this Unit is the statistical analysis of data. The purposes of data analysis are to characterize what is typical in the group, to indicate how widely individuals in the group vary, to show other aspects of how the individuals are distributed with respect to the variable being measured, to show the relation of the

different variables in the data to one another, and finally to describe the differences between two or more groups of individuals.

6.0   Exercise

Discuss the procedures for data analysis

What are the classificatory principles of categorization of data

Enumerate and discuss the reasons for statistical analysis of data

7.0   Further Readings and Other Resources

Selltiz, C., M. Johoda, M. Deutsch and S.W. Cook.
        Research Methods in Social Relations, Revised One-Volume Edition.
        London: Methuen & Co. Ltd. 1959

UNIT 12

BIOSTATISTICS: DATA PRESENTATION

Table of Content

1.0   Introduction

Unit 12 is a follow-up to Unit 11 that dealt with data analysis. After data have been collected and analysed, presention of data in the fom in which the users can understand becomes the next important step. This Unit therefore describes the use of figures, tables and graphs as ways in which data can be presented to the end users.

2.0   Objectives

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

Understand the use of figures in data presentation.

Explain the use of tables

Discuss the use of graphs

3.0 Use of figures

This is very common with quantitative data whose summaries given by calculating the measures of location and dispersion. The figures presented then are:

Arithmetic mean
Standard deviation
Size of sample

If the data were qualitative we simple quote the proportion with attribute in question its deviation and the total sample.

3.1 Use of tables

Data are presented in a precise form and this is mostly done by the use of tables. A table is the arrangement of data in horizontal rows and vertical columns. One major reason for tabulation is basically to reduce data based on variable, so as to give room for easy comparisons. Responses are reduced to sizeable number with the use of tables. A good table is expected to be comprehensive in communicating the contents to a lay reader. The comprehensiveness of a table does not mean that all data are presented in vague forms but a way of making the reduced data understandable.

Although there are no set rules in data tabulation, social scientist should bear in mind some hints necessary in the course of preparation of tables.

Tables should have precise and descriptive titles. The titles should be clear enough that a lay reader could easily grasp the summary of the contents of the tables.

There is the need for a layout of the tables in a rough sketch before actually drawing the tables. This is important so as to avoid rough drawing and ensure good tabulation.

The selected number of rows and columns must be given descriptive headings in which the quantities, variables and values are ascertained.

Delineate columns with the use of lines and border. Lines are appropriate when showing difference between headings, totals and other important aspects of the table from the table itself.

To avoid ambiguity in the interpretation of the table, the use of footnotes, asterisks and other symbols are important. This is necessary in such situations where the table emanates from more than one source.

Since neatness is part of academic integrity, always draw your table within a box frame for clarity and such should not be overcrowded.

An example of a table is presented below:

Infant and child Mortality Rates in Nigeria According to Major Characteristics, 1999.

Characteristics	Infant Mortality per 1000	Under-5 Mortality per 1000
Urban	59.3	107.8
Rural	75.4	142.9
No Education	76.9	156.8
Primary	71.2	121.8
Secondary	59.0	95.9
Higher	40.5	53.0
Northeast	79.4	175.2
Northwest	82.6	188.2
Southeast	74.3	135.0
Southwest	69.9	101.5
Central	50.7	84.4

Sources: Nigeria Demographic and Health Survey 1999, National Population Commission, Lagos, 200, P 100.

It may be observed in this table that the major conditions for the presentation of a table are taken into consideration.

3.2 Use of Graphs

Graphs are used to show a pictorial representation of data. The types of graph depend on the type of data available and the information needed to be highlighted. The major types of graphs are:
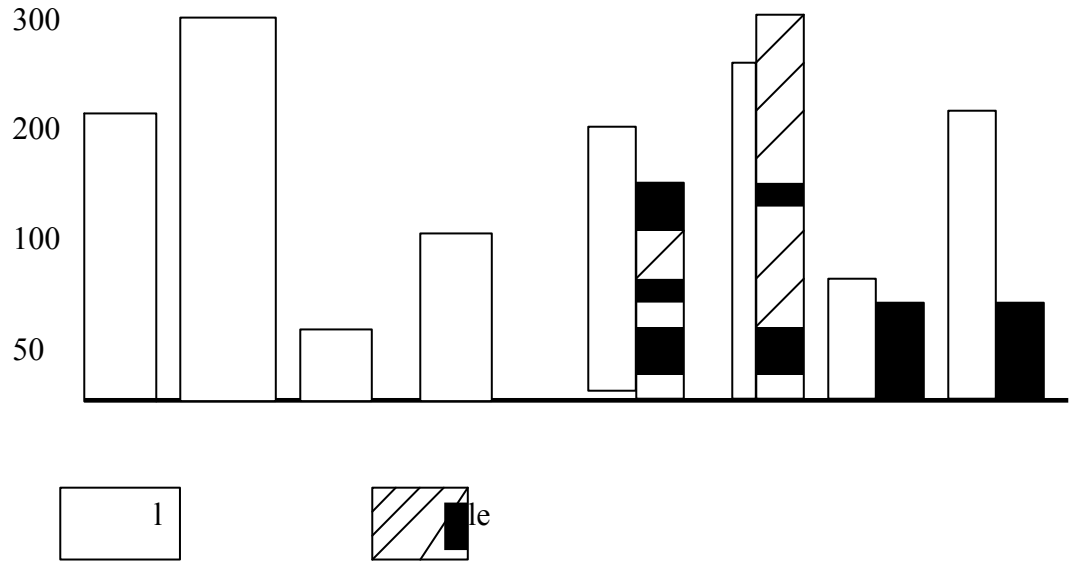
Bar charts

Histogram

Pie chart

These graphs like tables must have a good title and horizontal and vertical axis clearly labeled.

3.2.1 Bar Charts

These are drawn either horizontally or vertically to represents frequencies of observations. In some cases it is used to represent numerical observations but the base are separate and distinguished. It can be also used to represent percentage distribution of observations. The height corresponds with the frequency of observations of a category.

Simple Bar Chart   Multiple   Bar Chart

1		le	

Apart from simple and multiple bar charts, composite bar chart are also constructed to show the components of a whole. This shows the proportions of various sub-sections of the total observations. The length of the bar corresponds to this total and this is divided into parts that are proportional to be quantities that are being represented.

3.2.2 Histogram

histograms are useful for continuous variable and they are constructed by drawing small rectangles that adjoins each other with the area of the rectangles corresponding with the frequency of observations having the range of values.

5

0

| | 10 | 20 | 30 | 40 | 50 | 60 |

Age in years
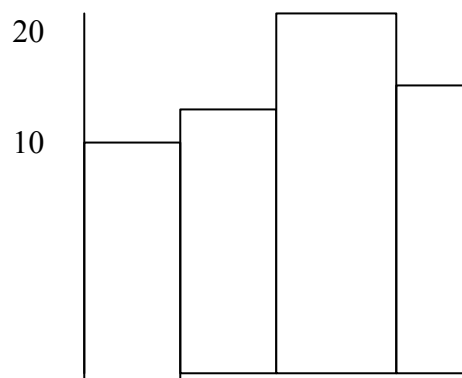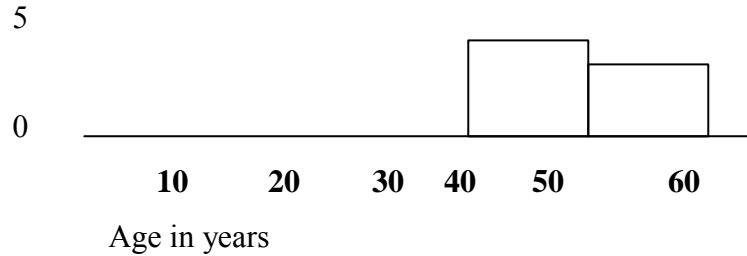
3.2.3.1 Pie Chart

Raw data can also be presented by the use of pie charts. Pie charts are constructed by drawing a circle and divided into segments proportionate to the size of each observation. The sum of angles of a circle at a center is 360^0 and this represents 100 percent of the total observation. This implies that 1 percent represent 3.6^0 for each category calculate the percentage proportional to the total and the corresponding degrees. Use a ruler and protractor to divide the circle into sectors proportional to the amount of the total, which they represent. An example of a pie chart showing the number of students in four departments in the faculty of social sciences in a Nigerian university in year 2002.

Departments	No of Students	Percentage Distribution	Central Angles
Economics	440	29.2	105.1^0
Political Science	435	33.3	119.9^0
Sociology	425	16.7	60.1^0
Geography	420	20.8	74.9^0
Total	1720	100.0	360.0^0

Number of students in four Department of a Faculty of Social Sciences of a Nigerian University in year 2002.

To obtain the percentages, divide each frequency of the 2nd column by their total (290), and multiply by 100. the central angles are obtain by multiplying the percentage by 3.6^0 . A pictorial representation of the above distribution of students in four departments of the Faculty of Social Sciences of a Nigerian University in year 2002 is shown below:

Pie Chart of Number of students in four Departments of the Faculty of Social Science of a Nigerian University in year 2002.
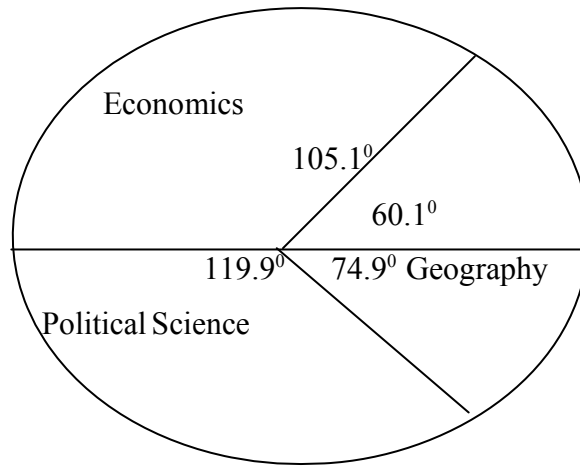
4.0 Conclusion

Data presentation is an essential component of research. The acceptability or otherwise of research findings depends on how the findings are presented to the end users. Data presentation therefore requires special skills . The format of data presentation also depends largely on the nature of the data, the use to which the data will be made and the audience.

5.0 Summary

Unit 12 discusses data presentation. It discussed the use of figures, tables and graphs as some of the ways in which qualitative data can be presented to the end users. The relative advantage of each of the methods of data presentation are also discussed.

6.0 Exercise

The number of currently married men and women who knew modern method of contraception in Nigeria in 1999 is presented below according to the five regions.

Region	Women	Men
Northeast	394	176
Northwest	386	186

Southeast	747	268
Southwest	1144	349
Central	894	277

Source: Nigeria Demographic and Health Survey 1999. National Population Commission, 2000, P. 46.

i.   present each of the findings in a simple bar chart

ii.   present the results for men and women in a multiple bar chart

iii.   construct a composite bar chart to show this information

iv.   present the same information using a pie diagram.

7.0   Further Readings and other Resources

Selltiz, C., M. Jahoda, M. Deutsch and S. W. Cook. Research Methods in Social Relations, Revised One-Volume Edition. London: Methuen *&* Co. Ltd. 1959.

UNIT 13

REPORT WRITING

Table of Contents

3.0   Introduction

3.1   statement of the Problem

3.2   The Research Procedure

3.3   The Results

4.0   Conclusion

5.0   Summary

6.0   Exercise

7.0   Further Reading and other Resources

1.0   Introduction

This is a follow-up to Unit 12 that dealt with data presentation. After data have been collected, presentation of data in the form in which the users can understand becomes the next important step. This Unit therefore describes what a research report should contain. A good research report should contain an introduction, statement of the problem, the research procedure, the results and the implications drawn from the results.

2.0   Objectives

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

Understand what a research report should contain.

Explain the processes in report writings.

Discuss the implications drawn from the results obtained from a study.

3.0   Introduction

The introduction deals with the general background information to the study being reported. It introduces the reader to the content of the main body. An introduction to a research report can be divided into sections. This depends on the researcher. The background knowledge and the focus of the report are contained in the introduction.

3.1 Statement of the problem

The statement of the problem serves to elaborate upon the information implied in the title of the study. The purpose of the study is set forth in brief but quite specific terms so that an uninformed reader may secure an idea of the study and manner in which it is to be conducted. In a nutshell, a research proposes to address a problem in a particular place and at a particular time. The problems that prompted the research are presented in clear statements so that the focus of the research can be grasp from the onset.

3.2 The research procedures

After setting out the background of the study and the statement of the problem, it is also necessary to present in clear and simple languages the method of data collection. Method of data collection depends on the nature and type of data. Primary data are pieces of information gathered for the sources at which they emanated while secondary data are data collected from organizations, establishments or ministries. The difference between the two is that while primary data is original, secondary data are already compiled by the organization involved but might not have been subjected to varying statistical analysis. In gathering data, it is necessary to understand the subject under study and how best they can be reached. For example, it may not be proper for a researcher to set forward his interview with prisoners on telephone. This is because prisoners may not have access to phoning facilities in the prison yard.

Among factors affecting the method of data collection is the literacy level of the respondents. It may amount to futility to send questionnaire to villagers to fill in response to some topical issues. Even when an interpreter is used, there is the tendency for misinterpreting the questions.

The period of data collection also matters. It may look foolish for a research wanting to interview rural people during the planting or harvesting season when almost all the people always go to farm. Accessibility to rural areas may be difficult most especially during the rainy season when the conditions of most rural roads are bad.

Generally speaking, research procedures deal with subject matter, their location, method and instrument of gathering them. The researcher should try to overcome all constraints in order to gather

data that are more reliable in generating ideas about the subject matter.

Method of gathering data may include the followings: random sampling, purposive sampling, cluster sampling, and stratified random sampling.   All these methods are such that a proportion of the total population is drawn out for data collection.   Instrument of data collection may include questionnaires, interview schedules, and tape recorder, notebook. The details are already discussed in Unit 10.

3.3   The results

The basic rule on presentation of findings is to give all evidences relevant to the research questions asked whether or not the results are in accordance with the researcher's views.   At the onset of any research outing, various objectives are set toward meeting these objectives and methodologies are also set. The results are presented to show the observation on the field. Observations from the field may or may not confirm the existing body of knowledge in the literature. In such a situation, new theoretical orientation emerges.

In order not to present vague results, data can be classified and presented in statistical forms. For example the data on the socio-demographic characteristics of a group of respondents may be presented in a table but such tables should not be overcrowded. The details have been discussed in Unit 12.

Other important aspects of presenting results are the clarity of expression and the use of simple language. Results of a research should be presented in such that a lay reader can grasp its content at first reading. As mentioned earlier, results should be drawn in line with research questions and objectives. They should be presented in the form of discussions, summary and conclusions. Based on the conclusion of a research work, recommendations can be made towards ameliorating the problem identified. For example, a social epidemiologist wanting to present a research report on the incidence of cholera outbreak in a community should be able to come with the recommendations that can help in containing the epidemic and avoid such occurrence in the future.

Research report cab be subjected to varying levels of statistical analysis such as descriptive statistics, measures of location, bivariate

and multivariate analysis depending on the nature of data and the researcher's orientation. If a study is policy oriented, results tell the reader what the various findings means in term of policy.

A research report must contain the summary of major findings and conclusions drawn from the findings. If the research is conducted to address a major problem in the community or the country or policy oriented it must include recommendations that are relevant to address the problem set out for investigation by the research. A good research report should also contain the list of references or bibliography consulted in the course of the research and in the preparation of the report. The references can be arranged in alphabetical or chronological order and attached to the end of the report usually before the Appendix.

Finally, it should be noted that it is not all the information that is related to a study that can be presented in the body of a report. Information that is more detailed and relevant are usually presented as Appendix to the Report.

4.0   Conclusion

A research effort is not complete until the report has been written. No matter how carefully designed and conducted a study is, the most interesting findings are not useful until they are communicated to others (in form of reports). Writing reports require special skills different from that required for the design, collections and analysis of data. When writing a report, it is necessary to know the purpose for which the research is being conducted. More often than not, the purpose of a research is to communicate with an audience. The audience might be less knowledgeable than the report. In such a situation, the researcher should use simple and clear languages in the presentation of a report to the end users.

5.0   Summary

This Unit, which is the final Unit in this discussed report writing. Report writing is an important aspect of any research and it requires a special skill. The research task is not completed until the report has been written. The carefully and scientifically conducted research has very little value unless the findings are communicated to the end users.

6.0   Exercises

Identify the major items, which a scientific report must contain.

Attempt a distinction between the contents of a research procedure and the contents of a research report.

7.0   Further Readings and other Resources

Selltiz, C., M. Jahoda, M. Deutsch and S. W. Cook. Research Methods in Social Relations, Revised One-Volume Edition. London: Methuen & Co. Ltd. 1959.

TutorMarkedAssignments

Q1   Define the basic concept statistics and distinguish between descriptive and inferential statistics?

Q2   Discuss the types of variables in social sciences and identify scales of measurement of qualitative and quantitative variables?

Q3   A group of 25 students obtain the following marks in Biostatistics:

58,34,57,54,44,57,61,21,36,57,45,47,38,48,51,60,61,65,51,75,58 54,51 45, 25.

(i)   Calculate the mean, median and the mode.
(ii)   Find the range, inter-quartile range, variance, standard deviation and the coefficients.
(iii)   Write a short note on your findings

Q4   Identify and discuss the various sources of statistics? Which of these sources are easily available in your country?

Q5   Discuss the major problems confronting successful conduct of population census in Nigeria.

Q6   Discuss ways in which population census and vital registration data can be used for planning in your community?

Q7   What are the goals of social and health research?

Q8   Explain the main features of exploratory, descriptive and experimental studies?

Q9   Identify and discuss the various stages of research?

Q10. What do you understand by sampling? Distinguish between probability and non-probability sampling technique?

Q11   Discuss the sampling methods that you will adopt in the study of use of health facilities in your community.

Q12   Enumerate the main methods of data collection. Explain the differences between open-ended questions and fixed alternative questions?

Q13   Discuss the procedure for data analysis.

Q14   Discuss four different ways in which data can be presented to the end users.

Q15   Identify and discuss the major items, which a scientific report should contain.