Lecture One

A. Basic Concept of Statistics

Definition of Statistics

Statistics is a mathematical science pertaining to collection, analysis, interpretation, and presentation of data. It is concerned with scientific method of collecting, organizing, summarizing, presenting and analyzing masses of numerical data so as to comprehend the essential features and relationship of the data. It can as well be defined as the study of the methods of collecting and analysis of data in such a way as to minimize any uncertainty in the conclusion drawn from the data. It assists in making decisions under uncertainties. Decision making process must be based on data neither on personal opinion nor on belief. A mass of data has little or no meaning until they are subjected to statistical analysis.

Major steps in statistical analysis are:

i. data collection
ii. Organizing and summarizing data
iii. Analyzing and interpreting the result
iv. Using the result to make rational decision.

B. Types of Statistics

There are two major classifications of statistics: descriptive and inferential statistics.

i. Descriptive statistics: This describes and summarizes the entire population through tables or charts, in order to bring out the important facts about the data. Examples of descriptive statistics are mean, median, mode and percentages.

ii. Inferential statistics: This is sometimes called objective or analytical statistics which is a method for studying only a part of the population in order to draw conclusions on the population based on the analysis of the sample data. Examples of inferential statistics are probability distributions, correlation and regression analysis. The main purpose of inferential statistics is to make an inference about a population based on the information contained in the sample.

C. Roles of Statistics

The following are some of the roles of statistics:

i. Statistics simplifies complex mass of data and presents them in a comprehensive way that they are at once made easy to comprehend and interpret. Instead of having a large raw data, the data are prepared in percentages and means which can be grasped more easily than a mass of data. Also, statistical analyses in the form of histogram, bar chart, or pie chart, make it easier for you to understand.

ii. Statistics presents data in more comprehensive and definite form: Statistics made conclusion that are stated numerically and more convincing than conclusions stated qualitatively. For

example, it is more attractive and convincing to say 85% of AGR 314 students passed in 2020/2021 session than saying most AGR 314 students passed in 2020/2021 session.

.iii. Statistics interpret conditions that are more presentable: Statistics present conditions in an attractive ways such as pie chart and histogram or bar charts of the phenomenon under investigation. Also, certain conditions are proved statistically to find out the probability of future occurrence of such situation so that necessary actions could be taken to prevent future occurrence of such conditions.

iv. It provides easy way of classifying numerical data: The method of classification in statistics provides the salient features of the variables that is under consideration. For example, statistical methods provide appropriate method of classifying two or more data by bringing out the maximum, minimum and the standard deviation of the various categories.

v. It provides an easy way of comparing data: Some data may be meaningless unless they are subjected to statistical analysis before they can be compared with similar data at other places. Statistics made an easy way of relating two different masses of numerical data by comparing some relevant information from the two data such as comparing their means, median and mode of their distribution

D. Statistics as a Subject and its Application

Statistics has very wide application in many fields. It is very useful in government‟s analysis and publication of many countries. For example, in population and housing, wages and salaries, budgets, education, agriculture, health, births and deaths of a country, etc. statistical tools are highly useful in population censuses and sample survey of national and international assignments.

E. Application Statistics in Different Fields of Study:

:i. Economics: Statistical data are highly useful in the understanding of economic policy and economic problems. In economics, numerical data are usually analysed statistically for ease of understanding. For example, volume of trade in Nigeria, price of commodities in certain year, wages and salaries of workers etc.

ii. Agricultural and biological sciences: In Agriculture, statistics is very important in crop experimental designs. It is used in comparing the results of crops from different fields and different designs. Statistical tools are also very useful in Biological Sciences as a means of testing the significance of results of experiments. It is also used in determining the relationship between the growth of different levels of feeding. It is applied in the estimation of fish population in a lake.

iii. Physical sciences: Some statistical methods were applied and developed in the field of physical sciences like Physics, Geology, Chemistry, etc. In the recent time, physical sciences are making good use of statistics in complex cases of molecular, nuclear and atomic structures.

F. Limitations of Statistics

Despite the usefulness of statistics in many fields, it also has some limitations and cannot be an answer to the whole affairs of the world. The following are some of its limitations:

i. Statistics deals with group or set of data and attach less importance to individual items. Statistics proves inadequate where the knowledge of individual items becomes necessary. It is most suited to those problems where aggregate characteristics are required.

ii. Statistics deals mainly with quantitative or numerical data: It is not all subjects that can be expressed numerically; there are situations where qualitative data are required. For example, poverty, intelligence and health are all qualitative data, which cannot be directly quantified. So, these types of data are not suitable for statistical analysis.

iii. Error during sampling could be to establish wrong conclusions if not handled by experts. Incorrect application of methods could lead to the drawing of wrong conclusions in statistics.

iv. Statistical data in most cases is usually approximated and not very exact. More emphases is usually laid on sampling method of data collection, that means that if a limited number of items are selected, the result of the sample may not be a true representation of the population.

G. Statistics and Logic

Statistics is the logic of common sense with arithmetic procedure. The logic supply the method by which data are to be collected and determined how extensive they are to be. The arithmetic together with certain readymade numerical tables yields the material on which to base inferences and measure the uncertainty of such inferences. The arithmetic is often routine and does not involve any special mathematics training. However, this is not to say that statistical theory is not from mathematics.

H. The Process of Statistical of Analysis

1. First describe and present your data, e.g. frequency distributions in tables or charts

2. Calculate basic statistics where possible, e.g. means and standard deviations

3. Start to interpret your data – what might it mean?

4. Select specific items for closer attention (based on your research hypotheses)

5. Select and carry out the right kind of test

6. Interpret your findings in terms of significance levels

7. Modify and repeat if necessary

I. Statistical Terms and Notations

- **Event**: an event is an outcome or defined collection of outcomes of a random experiment. Since the collection of all possible outcomes to a random experiment is called the sample space, another definition of **event** is any subset of a sample space.

- **Population**: A *population* is the entire group of individuals you want to study.
- **Parameter:** A *parameter* is a quantitative characteristic of the population that you're interested in estimating or testing (such as a population mean or proportion).
- **Sample** A subset of the population usually selected randomly. Measures that summarize a sample are called sample statistics.
- **A statistic** is a quantitative characteristic of a sample that often helps estimate or test the population parameter (such as a sample mean or proportion).
- **Measurement** is the assigning of numbers or codes according to prior-set rules. It is how we get the numbers upon which we perform statistical operations. Measurements that can vary or be expressed as more than one value throughout a study are called *variables*. For example, we may speak of the variable age, blood pressure, or height. In other words, variables represent the "thing" being measured.
- **Variable:** A variable is a characteristic of a unit being observed that may assume more than one of a set of values to which a numerical measure or a category from a classification can be assigned (e.g. income, age, weight, etc
- The **independent variable** of an analysis is the factor, intervention, or attribute that either defines groups or is thought to predict an outcome.
- The **dependent variable** is the measurement, outcome, or endpoint of the study.
- All variables other than the independent variable and dependent variable in a particular analysis are referred to as **extraneous variables**.
- **Continuous Variables**: These are variables which measure some count or quantity and don't have any boundaries are termed as continuous variables. It can be segregated into ratio or interval, or discrete variables. Interval variables have their centralized attribute, which is calibrated along with a range with some numerical values.
- **Discrete variables**: These are variables that assume only finite numbers number as values. i.e. distinct values with no intermediate points.
- **Categorical Variables:** It is a wide category of variable which is infinite and has no numerical data. These variables are called as qualitative variables or attribute variable in terms of statistics

J. Raw data

Raw data are unprocessed bits of information about some topic or event. The raw data need to be processed by subjecting it to some statistical analysis before it can make meaningful interpretation. A piece of data is called a score or observation while a collection of pieces of data pertaining to a topic is called variable. For example, the age of an individual is a score or observation, while the age of the entire group is a variable. Data can be collected from respondents through the use of structured questionnaires, group discussion (interview method), through phone or through email.

K. Classification of Data

Data could be classified as quantitative and qualitative. A quantitative data can be counted or measured e.g. age, income, years, farm size, inflation rate etc. Qualitative data on the other hand cannot be counted or measured but can only be explained, male or female, scale of operation, opinion of group of people etc. Due to the fact that statistics deals with numerical data,

qualitative data are usually converted into quantitative data by coding it or assigning value to it. For example, male = 1 and female = 0

Data could also be classified as primary and secondary data.
(i) Primary data are first-hand information. They are data collected by investigator himself or ask someone to collect for him, which have not been documented or analysed by someone else before. Primary data can be obtained through observation, interview method, questionnaire method, participatory rural appraisal method, etc.

(ii)Secondary data on the other hand is second hand information. It is an existing data already collected by someone else. An existing data may be in published or unpublished form. For example, in a study of the ages of graduating students of School of Agricultural Sciences, National Open University of Nigeria, the researcher may choose to observe and record it himself (primary data) or if the school keeps the records, he can use such data (secondary data).

Classification of Data

Data could also be classified as discrete and continuous data.

Discrete data have distinct value with no intermediate points. Examples of discrete data are household size, population census, number of students in a class, number of lecturers in a department, etc. The value of a discrete data must be in a whole number without fractions.

Continuous data can take any value. It could be a whole number or fraction. For example, continuous data are weight, age, height, yield of crops, etc