LESSON ONE: INTRODUCTION

1.1 Definition of Statistics

Statistics is a science that deals with the methods of collecting, organizing, presenting, analyzing and interpretation of numerical data to assist in making more effective decisions.

According to the above definition, there are five stages in a statistical investigation.

i) Collection

- It is the first step in a statistical investigation
- Data form the foundation of any statistical analysis and therefore should be collected with utmost care.
- If data are faulty, the conclusions drawn can never be reliable.

ii) Organization

- The large mass of figures that are collected from surveys usually need organization.
- The first step in organizing a mass of data is **editing** so that omissions, inconsistencies and irrelevant answers may be corrected.
- The next step is to **classify** some common characteristics possessed by the items constituting the data.
- The last step in organization is **tabulation**. The objective of tabulation is to arrange the data in columns and rows so that there is clarity

iii) Presentation

- After the data have been collected and organized they are ready for presentation.
- Data presented in an orderly manner facilitates statistical analysis

iv) Analysis

- It's a major step in any statistical investigation.
- Methods of analysis are numerous ranging from simple observation of data to highly mathematical techniques.
- We consider only the most common methods of statistical analysis

v) Interpretation

- It entails drawing conclusions from the data collected and analyzed.
- Correct interpretation will lead to valid conclusions of the study and thus can aid in decision making.

1.2 Types of Statistics

a) **Descriptive statistics**: It deals with processing data without attempting to draw any inferences from it. It refers to the presentation of data in the form of tables and graphs and to the description of some of its features such as averages.

b) **Inferential/Inductive statistics**: Refers to methods of using a sample to obtain information about a population i.e. making conclusions about the population based on information from the sample.

1.3 Population, Sample and Variables

- **Population**: is the totality of all the items or individuals whose characteristics we wish to study. Examples of a population are all the eligible voters in an election.
- **Sample**: is a subset or section of the population that is used to represent the whole population.
- **Parameter**: is any quantitative measure that describes a characteristic of a population e.g. population mean (μ) or population variance (δ^2).
- **Statistic**: is a quantitative measure that describes a characteristic of a sample e.g. sample mean ($\bar{x}$) or sample variance (s^2).

  E.G. The mean height of the people in Kenya is a parameter, whereas the mean height of a sample of 500 people is a statistic.
- Variable: is the characteristic that is being studied. It is represented by symbols *X, Y,* or *Z.* Height of people, grades in a test etc are examples of variables.
- There are two kinds of variables:

a) **Qualitative variables**: Are variables that are non-numeric i.e. attributes e.g. Gender, Religion, Color, State of birth etc.

b) **Quantitative variables**: are numeric variables e.g. the height of an individual when expressed in feet or inches, etc. Quantitative variables are either discrete or continuous.

   i) **Discrete variables**: Are variables, which can only assume certain values i.e. whole numbers. Are always counted. E.G: number of children in a family, the number of defective bulbs, etc.

   ii) **Continuous variables**: Are variables, which can assume any value within a specific range. Are always measured e.g. height, temperature, weight, radius etc.

1.4 Functions of Statistics

i) Definiteness i.e. statistics presents facts in a definite form:

- Statements or facts conveyed in exact quantitative terms are more convincing than vague utterances.
- Statements like "the population of Kenya is growing at a very fast rate", or "the prices of various commodities are rising", may not be very convincing as they don't specify the numerical dimensions involved.

ii) Condensation i.e. statistics simplifies a mass of figures

- Statistics helps in condensing a mass of figures into a few significant values e.g. mean, mode, median, standard deviation, etc.

iii) Comparison:

- Statistics facilitates comparison.
- Unless figures are compared with others of the same kind they are foten devoid of any meaning.

iv) It helps in formulation and testing hypothesis:

- Statistical methods are useful in formulating and testing hypothesis and to develop new theories.

v) Prediction and formulation of policies:

- Statistical methods provide useful means of forecasting future events.
- Knowledge of future trends is very helpful in framing suitable policies and plans.

1.5 Applications of Statistical Knowledge in Business Management

i) Marketing

- Statistical analysis are frequently used in providing information for marking decisions
- E.G: Analysis of data on population purchasing power, habits of people, completion, transportation costs etc should precede any attempt to establish a new market.

ii) Production

- The decision about what to produce, how to produce, when to produce, for whom to produce is based largely on facts analyzed statistically.

iii) Finance

- The finance mangers in discharging their finance functions efficiently depend heavily on statistical analysis of facts and figures.
- Financial forecasting, break even analysis and investment decisions under uncertainty are part of their activities.
- The area of security analysis is also highly quantitative.

iv) Banking

- Banks need to gather and analyze information on the general economic consideration.
- Banks' reserves are highly influenced by money markets which are not only local but also international.
- The credit department performs statistical analysis to determine how much credit to extend to various customers.

v) Purchase

- The purchasing department makes use of statistical data to frame suitable purchase policies such as where to buy, how to buy, at what time to buy and at what price to buy.

vi) Accounting

- The auditing function makes frequent applications of statistical sampling and estimation procedures.
- The account collects data on historical costs in the course of auditing a company's financial records and may use regression analysis to analyze the cost.

vii) Personnel

- The personnel department frames policies based on facts.
- It makes statistical studies of wage rates, incentive plans, cost of living, labor turnover rates, employment trends, accident rates employment grievances, performance appraisal, training programs etc.
- Such studies help the personnel department in the process of manpower planning.

viii) Investment

- Statistics greatly assists investors in making clear judgments in his investment decisions in selecting securities which are safe and which have the best prospects of yielding a good income.

1.6 Limitations of Decision-making

i) **Statistics does not deal with isolated measurement**

- Data are statistical when they relate to measurement of masses, not statistical when they relate to an individual item or event as a separate entity.
- E.G: The wage earned by an individual worker at any one time taken by itself is not statistical, but taken as a part of a mass of information, it may be a statistical data.

ii) **Statistics deals only with quantitative characteristics**

- Statements are numerical statements of facts. Thus qualitative characteristics like honesty, efficiency, intelligence etc cannot be studied directly.

iii) **Statistical results are true only on an average**

- The conclusions obtained statistically are not universally true; they are true only under certain conditions

iv) **Statistics is only a means**

- Statistical methods furnish only one method of studying a problem.
- They may not provide the best solution under all circumstances.
- Very often it may be necessary to supplement the conclusions arrived at by the help of statistical with other methods
- In deciding a course of action, it may be necessary to take into account other factors like the country's culture, religion, philosophy, personal, political or other non-quantitative considerations.
- Excessive dependence on statistics may lead to fallacious conclusions.

v) **Statistics can be misused**

- Statistics can be misused i.e. wrong interpretation. It requires experience and skill to draw sensible conclusions from the data.
- E.G: If statistical conclusions are based on incomplete information or there is bias in sampling.

1.7 Levels of Measurement

There are four levels of measurement; nominal, ordinal, interval and ratio.

a) Nominal scale

- It's the lowest level of measurement

- It merely groups observations into categories based on common characteristics eg gender, race, marital status, religion etc.
- Numbers are often assigned to the various categories for the purpose of identification. E.G: for the variable marital status we can assign 1 = married, 2 = single, 3 = divorced, 4 = windowed, 5 = separated.
- The numbers assigned to the various categories do no represent quantity or order and therefore performing mathematical operations on these numbers would yield meaningless values.
- The counting of members in each group is the only possible arithmetic operation when a nominal scale is employed. Accordingly we are restricted to use the **mode** as the measure of central tendency. There is mo measure of dispersion used for nominal scales.
- Chi-square test is the most common test of statistical significance.

b) Ordinal scale

- Items are not only grouped into categories but they are also ranked into some order. Therefore in an ordinal scale, numerals are used to represent relative position or order among the values of the variables.
- The use of ordinal scale implies a statement of 'greater than' or 'less than' (equality is also acceptable) without being able to state how much greater or less. The real difference between ranks 1 and 2 may be more or less than the difference between ranks 5 and 6.
- Since the numbers of this scale have only a rank meaning, the appropriate measure of central tendency is the **median**. A percentile or quartile measure is used for measuring dispersion.
- Correlations are restricted to various rank order methods. Measures of statistical significance are restricted to non-parametric methods.

c) Interval scale

- Numerals assigned to each measure are ranked in order and the intervals between them are equal. Hence numerals used represent quantity and some mathematical operations would yield meaningful values.
- However, the zero point is not meaningful, i.e. interval scales have an arbitrary zero and it is not possible to determine for them what may be called an absolute zero or the unique origin.
- The primary limitation of the interval scale is the lack of a true zero; it does not have the capacity to measure the complete absence of a trait or characteristic.
- The Fahrenheit scale is an example of an interval scale. One can say that an increase in temperature from 30^o to 40^o involves the same increase in temperature as an increase from 60^o to 70^o, but one cannot say that the temperature of 60^o is twice as warm as the temperature of 30^o because both numbers are dependent on the fact that then zero on the scale is set arbitrarily at the temperature of the freezing point of water. The ratio of the two temperatures, 30^o and 60^o, means nothing because zero is an arbitrary point.
- Intervals scales provide more powerful measurement than ordinal scales since the interval scale incorporates the concept of equality of interval.

- As such more powerful statistical measures can be used with interval scales. **Mean** is the appropriate measure of central tendency, while **standard deviation** is the most widely used measure of dispersion.
- Product moment correlation techniques are appropriate and the generally used tests for statistical significance are the**'t' test** and **'F' test**.

d) Ratio scale

- Ratio scales have an absolute or true zero of measurement. E.G: the zero point on a centimeter scale indicates the complete absence of length or height. But an absolute zero of temperature is theoretically unattainable and it remains a concept existing only in the scientist's mind.
- Ratio scale represents the actual amounts of variables. Measures of physical dimensions such as weight, height, distance, et. Are examples.
- All statistical techniques are usable with ratio scale and all mathematical operations (including multiplication and division) can be used
- Geometric and harmonic means can be used as measures of central tendency and coefficients of variation may also be calculated.