Introduction of Statistics

Bimal Shrestha

Lecturer

Nepal College of Information Technology

Balkumari, Lalitpur

Introduction to Statistics

- The word statistics is derived from the Latin word “status” or the Italian word “statista” or the German word “statistik” or the French word “statistique” each of which means the political state.

- In the ancient times the scope of the statistics was primarily limited to the collection of information related to the demographic information; and property and wealth of the country.

- Now, the statistics has great applications and play important role in any discipline such as economics, business, medicine, agriculture, etc.

Definition

- Statistics is the branch of science which is concerned with the collection, organization, summarization and analysis of the data (information); and drawing the inferences about the information collected during the investigation.

- Statistics is the science and art of collecting, summarizing, analyzing and interpreting the data that are subject to variation.

- Statistics is the study of methods and procedures for the collection, classification, analysis and interpretation of the quantitative data to make scientific inferences from it.

Statistics can be divided into two sub-categories:

1. Descriptive statistics, and
2. Inferential statistics

Descriptive Statistics

Deals with the collection, representation, calculation and summarization of the data to make it more informative and comprehensible.

Only few mathematical measures are used such as mean, median, standard deviation, range, etc.

Inferential Statistics

- Making inferences
- Hypothesis testing
- Determining the relationship
- Making prediction
- Concerned with the generalization of the results of a small sample to the larger population from which the sample is drawn.

Functions of Statistics

- Statistics collects the information from the scientific methods.
- It analyses the data.
- It simplifies the complexity.
- Inferences can be drawn from the analyzed data.
- Statistics presents the facts in a definite form.
- Statistics helps in formulation of suitable policies.
- Statistics facilitates comparison.
- Statistics helps in forecasting.

Scope of Statistics

- Statistics in Physical Sciences
- Statistics in Biological Sciences
- Statistics in Medical Sciences
- Statistics in Social Sciences
- Statistics in Business and Management Sciences
- Statistics in Information Technology

Limitation of Statistics

- Does not deal with individual data.
- Deals with quantitative data only.
- Statistical laws are not exact as mathematical laws. They are based on the average.
- One who is not expert in statistical methods gives poor and faulty conclusions.
- It sometimes gives absurd (difficult to understand) result.

Variables

Introduction

- A variable is any characteristics, number, or quantity that can be measured or counted. Variables are very common word in the statistics.
- The value of variable can vary from one entity to another. A variable is generally denoted by X, Y, etc.
- For example, the variables may be height, age, weight, blood pressure, pulse rate, blood sugar, blood group, disease, etc.
- A basic distinction in the nature between these variables is their quantitative or qualitative (categorical) measurements.

Quantitative Variable

- Quantitative variables are those characteristics which can be a count or measured numerically.
- Quantitative variable describes the characteristics in terms of numerical value, which are expressed in units of measurement.
- These type of data can be measured in the interval or ratio scale.
- Ex: height, weight, age, temperature, income, profit, etc.
- Quantitative variables can be divided into two types based of the nature of the characteristics.

i. Discrete Variable ii. Continuous Variable

Discrete Variable

- Discrete variable can take only specified number of values in a given range.
- In other words, a variable is said to be discrete if it takes only countably many values (whole number).
- For example, number of buses, family size, number of students, etc.
- Discrete variables are usually counts.

Continuous Variable

- Continuous variable can take any values over a particular range. It can assure either fractional or integral values.
- In other words, a variable is said to be continuous if it takes all possible real values (whole number as well as fractional values) within a certain range.
- Height of a person, weight, age, marks, blood pressure, temperature, are some of the examples for continuous variable.
- Here the obtained measurements can take any value in a given range.

Qualitative Variable

- Qualitative (categorical) variables are those characteristics which are not numerically measurable.
- These variables are either nominal (no natural ordering) or ordinal (ordered categories).
- It takes on values that are names or labels which have no mathematical meaning.
- The data are classified by counting the individuals having same characteristics or attribute and not by measurement.
- For examples: gender, disease, nationality, religion, etc.

Measurement and Scales

- Measurement is the process of assigning the numbers to objects with or without mathematical meaning according to certain rules. Measurement of statistical data is essential for further statistical analysis.
- Generally there are four types of measurement scales which are as follows.

1. Nominal scale
2. Ordinal scale
3. Interval scale
4. Ratio scale

Nominal Scale

- The lowest measurement scale is Nominal scale. As the name implies it consist of “naming” observations or classifying them into various categories.
- Sometimes the variable under the study is classified by the possession of some quality rather than by numerical measurement. In such case, the variables are measured using nominal scale of measurement .
- Some examples are gender : male/ female, religion, marital status, etc.
- Categorical data are generally measured on nominal scale.

Ordinal Scale

- Ordinal scale are slightly more advanced measure than nominal scale.
- All observations are not only different from category to category but can be ranked according to certain criterion then they are said to be measured on an ordinal scale.
- The label or category has been used in meaningful natural order but there is no information about the size of the interval.
- For example a patient's disease condition can be characterised as mild illness, moderate illness, or severe illness. The intelligence of children may be above average, average, below average. Some other examples are socioeconomic status, attitudes, pain, etc.

Interval Scale

- This scale of measurement is a more advanced scale than the nominal and ordinal scale where individuals are placed on a scale of continuous scale in which the distance between two measurement are well defined.
- Some measurement scales posses a constant interval size, thus they are called interval scale.
- The size of the interval between two observations can be measured.
- There is no true zero point but possesses only arbitrary zero. The addition and subtraction is used but the ratio of scores is not meaningful.
- For example the temperature of body.

Ratio Scale

- The highest level of measurement is the ratio scale.
- This scale is characterized by the fact that equality of ratios as well as equality of intervals may be determined.
- Fundamental to the ratio scale is a true zero point. Because of this reason ratio scale data are more meaningful.
- For examples height, weight, length, age, etc.. are measured using ratio scale .

Data Collection

Introduction

- The data can be defined as the facts or figures that can be expressed in numerical values and can be analysed further to get information about the population. The data are the qualitative or quantitative values of a variable. It is the lowest unit of the information. The data is the lowest unit of the information.
- The first step of research or any statistical study is the collection of data. Collection of data means the process that are to be used to gathering the information from the units under the study. The data are generally collected either from primary or secondary sources. The quality of data of information depends on the sources of data, methods of collection of data and recording system.

Primary Data

- The data collected by the investigator from the experimental studies for a specific research is called primary data.
- When the data is collected for the first time, data so obtained is termed as primary data.
- It is used in research when secondary data are not reliable and unavailable.
- These types of data are more reliable, accurate, and original in nature.
- They are usually obtained from the survey and enquiries conducted from the government, some individual researchers, and some research institutions.

Methods of Collecting Primary Data

- Personal interviews
- Indirect interviews
- Information from local reports
- Questionnaires
- Schedules
- Focus group discussion
- Case studies
- Observations checklist

Steps Followed in Collecting Primary Data

- Planning the study
- Preparation of schedules or questionnaire
- Selection of sample
- Editing the schedules

Preparation of Schedules and Questionnaire

There are no hard and fast rules for preparing questionnaire. However, the following are certain points which one should kept in mind while preparing questionnaire.

a. **Clarity** : The questions should be clear and simple. The ambiguous and confusing questions are avoided.
b. **Definiteness and objectivity** : The questions should be formulated in such a way that they have a definite and objective answer.
c. **Numbers and arrangement of questions** : As much as possible, the number of questions should be small and questions should be arranged coherently.

Continue…

d. Reliability and accuracy : Questions should be reliable. The level of accuracy should be also clearly stated.

Secondary Data

The data which are already collected by other agencies, or authorized institutions is known as secondary data. The reliability and validity of the data depends on the sources of data. This secondary data are not original in nature. There are some published or unpublished sources of secondary data such as government publication, national and international reports, also unpublished articles, thesis, etc.

The secondary data have some advantages over primary data

1. Highly convenient to use information
2. Easily available and obtained quickly
3. Possible to obtained in large volumes and scales.

Continue..

However, these are not sufficient to meet the need of present researches and those data may not be accurate and adequate for the purpose of further research. The idea for selecting the primary or secondary data is given by the following points.

- The nature and scope of inquiry
- Availability of finance and human resources for the desired data
- Availability of time
- The degree of adequacy and accuracy of desired data
- The data collecting agency like government, institutions, individuals, etc.

Sources of Secondary Data

- Published sources:

i. Official publications of different ministries and department of government e.g. National planning commission, CBS, etc.

ii. Reports and publications of organizations like NHRC, NMC , etc.

iii. Reports of economists, research scholars, universities etc.

iv. Official publications of INGOs and NGOs

Continue..

- Unpublished sources:

i. Records maintained by government offices

ii. Records maintained by research institutes, research scholars, etc.

iii. Records updated by the institutions for their internal purpose.

Thank You !!!