

Introduction to Statistics

Statistics:

In singular form,

Statistics is the science of collecting, presenting, analyzing and interpreting data to assist in making more effective decision.

In plural form,

A collection of numerical information is called statistics.

Example:

Life expectancy of Bangladesh is 70.65

Types of statistics:

There are two types of statistics

1. Descriptive statistics
2. Inferential statistics.

1. Descriptive Statistics: Descriptive statistics involve graphical and numerical procedures that are used to summarize and process data and to transform data into information.

Example:

A study found that 49% of the people in a survey knew the name of the first book of the Bible. The statistic 49 describes the number out of every 100 persons who knew the answer.

2. Inferential Statistics: inferential statistics are used to make inferences about a population from a sample in order to generalize (make assumptions about this wider population) and/or make predictions about the future. Inferential statistics provide the bases for prediction, forecasts and estimates that are used to transform information into knowledge.

Example:

TV networks constantly monitor the popularity of their programs by hiring Nielsen and other organizations to sample the preferences of TV viewer.

***Descriptive statistics is the discipline of quantitatively describing the main features of a collection of information, or the quantitative description itself. Descriptive statistics are distinguished from inferential statistics (or inductive statistics), in that descriptive statistics aim to summarize a sample, rather than use the data to learn about the population that the sample of data is thought to represent.

Some important definitions:

Population: A Population is a Collection of all possible individuals, objects, or measurements of interest.

Sample: A sample is a representative part of the population.

Parameter: A Parameter is a measurable characteristic of a population.

Statistic: A statistic is a measurable characteristic of a sample.

Application areas: modeling of text and Web data, network traffic modeling, probabilistic analysis of algorithms and graphs, reliability modeling, simulation algorithms, data mining, and speech recognition, Machine Learning portion, data science etc.

Limitations of Statistics

The limitations of statistics are being pointed out as follows:

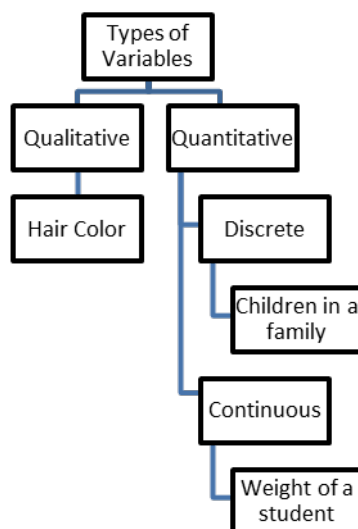
- ⊙ Statistics does not deal with qualitative phenomenon. However, qualitative phenomenon can be analyzed if it is expressed numerically or if it is scored after assigning appropriate number.
- ⊙ Statistics deals with population and statistical methods deal with a mass of data and not with a single figure.
- ⊙ Statistics can be misused. The data placed in the hands of an inexperienced person may lead to be compiled inaccurately and interpreted unscientifically.
- ⊙ Statistical results are true only on an average.

Variable

Any characteristics of a population which may vary from person to person or object to object either in magnitude or in quality are called variables.

Example

- a. The heights of adult males
- b. Sex of a person



Qualitative Variable or an attribute

The variable for which numerical measurement is not possible and has only several mutually exclusive categories is called qualitative variable.

Examples

- a. Gender (Male, Female)
- b. Religion (Muslim, Non-Muslim),
- c. Eye color (Blue, Black, Brown)
- d. Marital Status (Single, Married, Widowed, Divorced, Separated)

Quantitative Variable

The variable for which numerical measurement is possible and thus possesses a natural ordering is called quantitative variable.

Examples

- a. Height of a person
- b. Weight of a person
- c. family size
- d. No of accident in one day etc.

Quantitative variables are either discrete or continuous.

i. Discrete Variable

When a variable can assume only the isolated values which are countable within a given range is called discrete variable.

Examples

- a. No of children in a family
- b. Family size etc.

ii. Continuous Variable

When a variable can assume any value within a specific range is called continuous variable.

Examples

- a. Height of a person
- b. Weight of a person etc.

Levels of Measurement

There are actually four levels of measurement.

1. Nominal-Level Data
2. Ordinal-Level Data
3. Interval-Level Data
4. Ratio-Level Data.

1. Nominal-Level Data:

- In nominal data, the categories of variables differ only one another in name only.
- In other words, one category of a variable is not necessarily higher or lower, or greater or smaller than the other category.
- Even when the labels are numerically coded, the data categories have no logical order.

Example:

- a. Gender
- b. Religion (Muslim, Non-Muslim),
- c. Eye color (Blue, Black, Brown)
- d. Marital Status (Single, Married, Widowed, Divorced, Separated)

2. Ordinal –Level Data:

- There is an ordered relationship among the categories.
- You can level/rank your data.

Example:

- a. Your level of knowledge in Statistics (good, average or poor)
- b. Your visit in the hospital (frequent, occasional, rare and never)

3. Interval –Level Data:

- An interval variable is one where the distance (or interval) between any two category is known and constant.
- An interval level data is that it does not have a true zero point.

Example:

- a. Temperature:

It is approximate to say that the difference in the temperature between two objects, one having temperature 20°C and the other having temperature 60°C is 40°C , but it is not correct to say that the object having 60°C is thrice as hot as the other object having temperature 20°C .

- b. IQ test score.

4. Ratio –Level Data:

- Meaningful zero point and ration between variables, e.g. the zero point is the absence of the characteristics and the ratio between two numbers is also meaningful.

Example:

- a. Height
- b. Weight
- c. Distance