

Introduction to Statistics

History of statistics

Originally the word statistics has been derived from

1. The Latin word “status”
2. The Italian word “statista”
3. The German word “statistik”
4. The French word “statistique”

All of these four words have same meaning” political state”

Gottfried Achenwall has first time used the word statistik during one of his lecture at a German university.

Shakespeare and Milton have also used the word “statist” in their plays.

Definitions of Statistics

1. Statistics is the collection, presentation, analysis and utilization of numerical data to make inferences and reach decisions in the face of uncertainty in economics, business and other social and physical sciences.
2. A systematic scientific method executed for analysis or numerical data with the object Of estimation, interpretation and/or prediction.
3. Statistics is the science of problems-solving in the presence of variability.
4. The scientific study of techniques for collecting, analyzing, and drawing conclusion From data.
5. Statistics is the science learning from data.
6. It is science that deals with the collection, classification, analysis and making Inferences from data or information

Main Branches of Statistics

The science of statistics may be classified into two main branches namely:

- a) Theoretical/Mathematical/Pure statistics
- b) Applied Statistics

a) Theoretical statistics.

It is a branch of statistics that formulates statistical methods and general rules to be applied for investigation to a specific problem.

This branch of statistics are divided into three branches

i) Design of experiment.

The design of experiment is a sequence of steps taken to collect appropriate data for objective analysis to draw valid inference with respect to a problem under investigation.

ii) Descriptive (or deductive) Statistics.

A branch of statistics in which we collect, arrange and analyze the numerical data and no conclusion is drawn about the population is called descriptive statistics.

Example:

If we select 10 students of first year class of a college and find their mean age. Let the mean age is 16 years. Now if we interpret our result as that mean age of these 10 students is 16 years then this is descriptive statistics.

iii) Statistical inference (or inductive) statistics

A branch of statistics in which we collect, arrange and analyze the numerical data and also draw conclusion about the population is called inferential statistics.

Or

This branch of statistics deals with drawing conclusions about population on the basis of the sample information.

In statistics the inferences are mostly drawn in two ways;

1. About population parameters: Here we draw inferences on the basis of sample
2. About uncertainty: we are attaching some confidence or certainty to uncertain events with the help of probability theory.

Example:

If we select 10 students of first year class of a college and find their mean age. Let the mean age is 16 years. Now if we conclude that mean age of all first year students is 16 years then this is inferential/inductive statistics.

b) Applied Statistics

Applied statistics is a branch of statistics that makes use of statistical methods and general rules in the investigation of a specific problem. This branch of statistics is applied in the fields of agriculture, industry, transport, business, banking, insurance, economics and social sciences

Meaning of statistics

At present statistics is used to give the following three meanings:

a) In plural sense, statistics means data, systematically collected and presented in the form of tables or charts etc.

Examples:

1. Statistics of prices
2. Statistics of road accidents
3. Statistics of deaths
4. Statistics of births, etc

b) In singular sense, statistics means methods. These methods include collection, tabulation, presentation, analysis and interpreting of numerical data.

Examples:

1. Marks obtained by students in Mathematics and Physics
2. Heights of fathers and sons. Etc

c) In plural sense Statistics means statistic:

Or

Statistic means numerical quantity calculated from sample values.

In this sense the word use as plural.

Example:

If we select at random 10 students from a class of 50 and we want to measure their heights and average of heights it is statistics.

Population

A set of all possible observation whether finite or infinite, relevant to characteristic of interest is called population.

Or

Total group under discussion is called population.

Examples:

1. The population of stars on the sky.
2. The number of books in the library.
3. The number of students in the class.

Sample

A representative subset of a population is called sample.

Or

A part selected from population is called sample.

Examples

1. The few drops of blood taken from a human body.
2. Few plants selected from a standing crop for research purposes
3. A small amount of soil to be taken for analysis in laboratory

Parameter

A numerical value calculated from the population is called a parameter

Examples: $\mu = \frac{\sum X}{N}$ $\sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}}$

Statistic

A numerical value calculated from the sample is called a statistic

Examples: $\bar{X} = \frac{\sum X}{n}$ $S = \sqrt{\frac{\sum (X - \bar{X})^2}{n}}$

Variable

A characteristic that varies from one individual to another individual or object

Example

Age, height, prizes

Types of variable

- a) Qualitative variable
- b) Quantitative variable

Qualitative variable

A Variable/characteristics which changes only in quality from one individual to another individual and cannot be measurable is called qualitative variable. Qualitative variable is also called attribute.

Examples

Eye, hair colour, kindness, sex and beauty

Quantitative variable

A Variable/characteristics which changes only in quantity from one individual to another individual and can be measurable is called quantitative variable.

Examples

Height, weight, length and prices

Random variable

A variable which is unpredictable is called random variable

Example

Scores of batsman, outcomes of a die

Constant

A characteristic which does not vary but remains fixed is called a constant

Example

$\pi = 22/7 = 3.14159$

Discrete variable

A countable variable is called as discrete variable

Such variable takes values as whole integers

e.g.: 0, 1, 2,n

- 1) No. of children in a family
- 2) No. of eggs laid by a hen in a year
- 3) No. of tosses of a coin
- 4) No. of student in a class

Discrete variable is also called discontinuous variable. It does not take fractional values i.e. 0.333, 1.33, 25.26 etc

Properties

- 1) The values of discrete variable are countable
- 2) It assumes finite number of values in the given interval
- 3) The graph of discrete variable is discontinuous with breaks and each point touches in the form of lines at x-axis

Continuous random variable

A measurable variable is called continuous variable. Continuous variable takes without any gap. It takes values within two limits.

Continuous variable takes values as fractional, decimal, integral, whole numbers etc

Examples

1. Speed of a car
2. Age, height, weight
3. Distance, length

Properties

1. It assumes all possible values without jumping on the measuring scale
2. It assumes values in whole numbers and in fractions measured to any numbers of decimal place
3. It assumes infinite number of values in the given interval
4. The graph of continuous variable is a continuous curve without any breaks

Characteristics of Statistics

The word Statistics in plural sense has some characteristics which are discussed below.

1. Statistics are aggregate of facts.
2. Statistics are numerically expressed.
3. Statistics deals with quantitative data only but qualitative data such as honesty, beauty, poverty, liking, intelligence, etc cannot be studied directly.
4. Statistics are collected in a systematic manner.
5. Statistics are collected for a predetermined purpose.
6. Statistics must be comparable with each other.
7. It tests the laws of other sciences.
8. Statistical test are true for the long run.
9. Statistics provides only tools for analysis.
10. Statistics are estimated according to a reasonable standard of accuracy.

Importance of statistics in different fields

Statistics has come to play an important role in every field of life.

1. It plays an important role in business
2. all types of banks make use of statistics for a number of purposes
3. the whole structure of insurance is based on statistical data
4. Statistics has an important position in almost all the natural and social sciences.
5. statistics has prove to be of immense use in physics and chemistry
6. astronomy is one of the oldest branches of the statistical study
7. statistics has prove to be of immense use in meteorology because statistical principals and method are use in weather forecasting
8. Statistical data are widely use in biology for measurement of living organisms like human beings, animals and plants etc.
9. it plays an important role in economics
10. the role of statistics in the subject of education, psychology and sociology cannot be ignored

Limitations of statistics

Some Limitations of statistics are listed below

1. Statistical laws are true on the average.
2. Statistics deals with aggregates.
3. Statistics cannot be applied to heterogeneous data.
4. One of the greatest limitations of statistics is that it deals with the characteristics which can be numerically specified.
5. If sufficient care is not exercised in collecting, analyzing and interpreting the data, statistical result might be misleading.
6. Only a person who has an expert knowledge of statistical methods can handle statistical data.

Functions or Uses of Statistics

1. Statistics simplifies complexities.
2. Statistics presents facts in a definite form.
3. Statistics simplifies comparison of data.
4. Statistics studies relationship among different facts.
5. Statistics studies change in the level of a given phenomenon.
6. Statistics aids forecasting.
7. Statistics guides the formulation of policies.
8. Statistics tests the laws of other sciences.

Level of measurement

If a variable in mathematics and statistics is a classification that is used to describe the nature of information contained within numbers assigned to objects and therefore within a variable

There are four types

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

Nominal measurement

In this type of measurement names are assigned to objects as labels.

Examples

1. Your social security numbers
2. Town house
3. Yes or No
4. Male/Female
5. Telephone number

Ordinal measurement

In this classification the number are assigned to objects represent the rank order

Examples

1. Level of agreement: No; may be; yes.
2. Political agreement: left; centre, right.
3. Years in college.
4. Social sciences.
5. 1st, 2nd, 3rd place in a contest.

Interval measurement

The numbers assigned to objects have all the feature of ordinal measurement and in addition equal difference between measurements represent equivalent

Examples

1. Measurement of sea level.
2. Income
3. Temperature using thermometer.
4. Standardized test scores i.e. Sat and GRE
5. Time of day as 12 hour clock.

Ratio measurement

A ratio measurement scale is one in which the ratio between any two measurements is meaningful.

Examples

1. No. of clients in past six month.
2. No. of organizations belongs to.
3. No. of Question asks during the lecture.
4. No. of risk to build software.
5. No. of children.