

Chapter - 1

Introduction of Statistics

1. Origin and development of Statistics

Statistics is an old science, originated during the time of Mahabharat. The word 'Statistics' appear to have been derived from

Latin word 'Status.'

Italian word 'Statista'

German word 'Statistik'

French word 'Statistique.'

In ancient time, the government used to collect the information regarding the population and property of wealth of the country.

For the last few centuries, Mathematicians like Pascal, James Bernoulli, De Moivre, Laplace, Gauss, Lagrange, Bayes, Markoff, Euler etc were mainly interested in the development of the theory of probability as applied to the theory of games and other chance phenomena. Till the early nineteenth century, statistics was mainly concerned

with official statistics needed for the collection of information on revenue population and area of land under cultivation.

The scope of statistics has been developing gradually and its field of application has been increasing day by day.

1.2 Definition of Statistics

The process of planning experiments, obtaining data and then organizing, summarizing, presenting, analysing, interpreting and drawing conclusion based on the collected data is known as statistics.

or

Statistics is the science which deals with collection, classification and tabulation of numerical facts as the basis for explanation, description and comparison of phenomenon.

1.3 Importance and Scope of Statistics

- statistics in planning.
- statistics in state.
- statistics in mathematics.
- statistics in economics

- statistics in business and management.
- statistics in industry.
- statistics in insurance.
- statistics in astronomy
- statistics in war
- statistics in engineering
- statistics in biology and medical science.

1.4 Limitation of statistics

i) Statistics is not suited to the study of qualitative phenomenon.

Statistics is not applicable to study qualitative phenomena like honesty, happiness, sorrow, poverty, culture etc which cannot be expressed numerically are not capable of direct analysis.

ii) Statistics does not study individuals.

Statistics deals with an aggregate of objects and does not give any specific recognition to the individual of a series.

iii) Statistical laws are not exact.

Statistical laws are only approximation and not exact. Like physical and natural sciences. On the basis of statistical analysis we can talk only in terms of probability and chance and not in terms of certainty.

iv) Statistics is liable to be misused.

Statistical methods are most dangerous tools in the hand of the clumsy/unskilled/inexpert. The requirement of experience and skill for judicious use of statistical method restrict their use to experts only and limit the chances of the mass popularity of this useful and important science.

Unit - 2

Collection of data

Data

In statistics, data refers to facts, numbers or information collected, observed or recorded for analysis. It can be in the form of measurements, observations or descriptions of things and it serves as the basis for making conclusions, predictions or decisions.

Types of data

Statistics is the collection of information about the concern study, mostly in the form of data.

The types of data collection are:-

- i) Primary data
- ii) Secondary data.

i) Primary Data

Primary data refers to original data collected first hand from its source.

This information is gathered directly by researchers. The data obtained in a census study are termed as primary data.

Methods of collecting primary data.

The methods of collecting primary data are as follows:

- Surveys

- Interview

- Observations

- Experiments

- Focus group

- Case study

- etc

P₂ Secondary data

Secondary data refers to information that has been previously gathered, collected, and recorded by someone else for a different purpose than the one you might have in mind.

This data can be obtained from various sources, such as books, journals, articles, etc.

For example, the data given in different census years is again processed to obtain trends of population growth, mortality rate, death rate, etc.

Sources of Secondary data.

The sources of data can be (are):-

- Government Sources
 - Academic Institutions
 - Commercial Source
 - Non-Governmental Organization (NGOs)
 - Media
 - Online platforms
 - Books
 - Journal
 - Publications, etc.
-

Methods of collecting primary data

i) Surveys

These involve structured set of questions given to respondents to gather their opinions, behaviour or demographics.

ii) Interviews.

Conducting one-on-one or group discussions with individuals to gather detailed insights, opinions or experiences.

iii) Observations

Directly observing and recording behaviour, events, or activities without interfering in the natural setting.

iv) Experiments

Manipulating variables in a controlled environment to observe the effects and gather data.

v) Focus group

Small group discussions moderated by a facilitator to explore perceptions, attitudes or opinion on a specific topic.

vi) Case Study

In-depth analysis of a particular subject, often used in social sciences or business studies.

Unit - 3

Classification and Tabulation

Classification

Classification is a way of sorting things into different groups or categories based on their similarities or characteristics they share. It helps to organize and understand information by grouping.

Or,

The placement of data in different homogeneous groups, formed on the basis of some characteristics or criteria is called classification of data.

Tabulation

Tabulation refers to the systematic arrangement of data in a table or tabular form, usually presenting information in rows and columns for easy comprehension, comparison, or analysis. It organizes data in a structured format, making it easier to understand patterns, relationships, and summaries within the data set.

Objects of classification

The main objectives of classification of data are:-

- It presents the facts into a simpler form.
- It condenses the huge mass of data by eliminating irrelevant details such that point of similarities and dis-similarities can be brought out.
- It facilitates comparison.
- It prepares the basis of tabulation.

Classification of data

Data classification is the process of organizing and categorizing data into different groups and classes based on certain criteria or characteristics.

Common types of data classification are as follows:

1. Geographical classification:

In geographical classification, data are arranged according to places, areas or regions.

ii) Chronological classification

Chronological classification typically refers to the arrangement of things, events or information in a sequence based on their time of occurrence or creation.

iii) Qualitative classification

Qualitative classification involves categorizing data into descriptive groups or classes based on non-numeric characteristics or attributes. It focuses on characteristics that cannot be measured numerically.

iv) Quantitative classification

Quantitative classification is a statistical technique to classify data based on numerical or quantitative attributes. The height, weight, income of person, etc are the examples which can be measured quantitatively.

Importance of table

Tables in statistics are fundamental tools used to organize, summarize and present data in a structured format. Some of the importance of table are:-

i) Data organization

Tables systematically arrange raw data, making it easier to interpret and analyze.

ii) Summary and Comparison

Tables allow for the summarization of data through calculations like totals, average, percentages, etc.

They enable comparisons between different categories or groups, facilitating a better understanding of variations and differences.

iii) Clarity and Communication

Tables present information in a concise and easy-to-understand format, enhancing the clarity of data presentation.

iv) Facilitation of Statistical Analysis

Tables serve as the basis for various statistical analysis, such as calculating means, medians, standard deviations, correlations and more.

v) Visual Representation

Tables offer a visual representation of data, making it easier to convey information, statistics or comparisons to an audience.
etc.

Parts of table

- i) Title : A descriptive title that summarizes the content or purpose of the table.
- ii) Column headings : Names or labels for each column, representing the variables or categories being measured.
- iii) Row headings : Names or labels for each row, indicating different cases, group or observations.
- iv) Body : The main section of the table containing the data organized into rows and columns.
- v) Footnotes : Explanatory notes or additional information relevant to the data in the table.
- vi) Sources :
These parts collectively help in presenting and understanding the information in a structured and organized manner.

For example:

Table no 3.1

Head notes

Sub	Boys		Girls		Total
	Column	column	column	column	
	Head	Head	Head	Head	
	→		Body	←	

Foot notes

Source:

Frequency

Frequency is the number of times a particular value or category occur. Example, if the variate value x repeats 5 times, then the frequency of x is 5.

Frequency distribution

Frequency distribution is simply a table in which the collected data are classified and presented in different groups. There are two types of frequency distribution

A) Univariate Frequency Distribution

i) Individual Series

Individual series is a series where items are listed singly after observation as distinguished from listing them in group.

For example:

Day	Sun	Mon	Tue	Wed	Thurs	Fri	Sat
Temp(°C)	32	29	27	30	34	28	31

ii) Discrete Series

The distribution which can be formed by a discrete variable is known as discrete series.

For example

No. of goals	0	1	2	3	4
No of Match	27	9	8	5	2

iii) Continuous Series

The series formed by continuous variable is called continuous series

For example

Wages	100-200	200-300	300-400	400-500
Frequency	25	16	7	3

B = Bivariate Frequency distribution

The distribution which has two variables under the study is called bivariate frequency distribution.

For example:

- i) Income and expenditure of a group of individuals
- ii) Height and weight of a group of persons
- iii) Age and height of a group of boys.

Examples

- 1) The marks obtained by 30 students of a class are given below:

40 60 70 30 38 25 42 50 58 59 46 77 50
54 35 42 65 72 34 20 35 48 54 53 60
78 62 51 42 28

Classify the above data taking a class interval of 10.

Class Interval	Tally Mark	Frequency
20-30		3
30-40		5
40-50		6
50-60		8
60-70		4
70-80		4

2. Construct the discrete frequency distribution from the following data

Mark: 50, 67, 71, 50, 67, 79, 50, 61, 71,
50, 80, 80, 79, 71, 50

Mark	Tally Mark	Frequency
50		5
61		1
67		2
71		3
79		2
80		2