

1

Introduction to Statistics

Important terms and concepts

- ❖ **Statistics** simply means numerical data, and is field of math that generally deals with collection of data, tabulation, and interpretation of numerical data. It is actually a form of mathematical analysis that uses different quantitative models to produce a set of experimental data or studies of real life. It is an area of applied mathematics concern with data collection analysis, interpretation, and presentation. Statistics deals with how data can be used to solve complex problems. Some people consider statistics to be a distinct mathematical science rather than a branch of mathematics.
- ❖ **Economics** is a science that studies human behavior which aims at allocation of scarce resources in such a way that consumer can maximise their satisfaction, producers can maximise their profits and society can maximise its social welfare. It is about making choice in the presence of scarcity.
- ❖ **Scarcity** means shortage of goods and resources in relation to their demand.

Scarcity is the root of all Economic problem -

- ❖ **Father of Economics Adam Smith provided wealth definition of economics(Book- The Wealth of Nations, 1776):-** Economics is an enquiry into the factors that determine the wealth of a country.
- ❖ **Scarcity definition given by Lionel Robbins (1932):-** Economics is a science that studies human behaviour as a relationship between ends and scarce means which have the alternative uses.
- ❖ **Growth definition given by Paul A Samuelson:-** Economics is the study of how man and the society chooses with or without the use of money, to employ scarce productive resources which could have the alternative uses to provide various commodities overtime and distribute them for consumption now and in future among various people and the groups of the society.

Resources are:

1. Scarce/limited and
2. Have alternatives uses

Types of Activities -

- ❖ **Economic Activities**
 - Production
 - Consumption
 - Investment
 - Exchange
 - Distribution
- ❖ **Non-Economics Activities**
 - Social
 - Religious



- Political
- Charitable
- Parental
- ❖ **Economic activities** are those activities which are related to earn money and wealth for life. These activities generate new income and increase the flow of goods and services. For example production, consumption, investment, distribution.
- ❖ **Non-economic activities** are those activities which are not related to earn money and wealth. These activities neither generate income nor increase the flow of goods & services. For example, a teacher teaching his own son.
- ❖ **Consumer:** Consumer is an economic agent who buys the goods and services to satisfy his wants.
- ❖ **Producer** is one who produces goods and services for the generation of income.
- ❖ **Service holder:** A person who is in job and gives his services as a factor of production and is getting paid for it. E.g. Govt. Teacher.
- ❖ **Service Provider:** A person who provides services to other for a payment. e.g. transporter, auto driver.
- ❖ **Statistics:** Statistics is a method of taking decisions on the basis of numerical data.

Statistics can be defined in two ways -

- ❖ **Singular sense:** Statistic Means Statistical methods and techniques related to collection, organisation, classification, Presentation, analysis and Interpretation of data.
- ❖ **Plural Sense:** Statistics means Numerical facts and figures Which have been systematically collected for a definite purpose in any field of study.

Characteristics of statistics in plural sense -

- ❖ Aggregate of facts
- ❖ Numerically expressed
- ❖ Affected by multiplicity of causes
- ❖ Reasonable accuracy
- ❖ Collected in a systematic manner
- ❖ Pre-determined Purpose
- ❖ Placed in relation to each other

Stages of statistics in singular sense -

- ❖ Collection of data
- ❖ Organisation of data
- ❖ Presentation of data
- ❖ Analysis of data
- ❖ Interpretation of data

Statistical data -

- ❖ **Qualitative data:** Not measured in numerical terms like beauty and intelligence.
- ❖ **Quantitative data:** Measured in numerical terms like price and Income.

Scope of Statistics -

In the old days, the use of statistics was restricted to deal with the affairs of the state. But now-a-days the scope of statistics has spread to all those areas where numerical facts are used such as economics, business industry, medicine, physics, chemistry and numerous other fields of knowledge.

Importance of Statistics in Economics -

- ❖ It enables an economist to present economic facts in precise and definite form.
- ❖ Helps in condensing mass data into a few numerical measures.
- ❖ Statistics is used in finding relationship between different economic factors.
- ❖ Economics forecasting through statistical studies.
- ❖ Helpful to formulate appropriate economic policies that solve economic problems.
- ❖ Help to analyse the performance of policies applied before.
- ❖ Economist try to find out cause and effect relationship between different sets of data.
- ❖ Formulation of policies.
- ❖ Used for inter-sectoral and inter-temporal comparisons.
- ❖ It is a quantitative expression for economic problem.

Function of Statistics

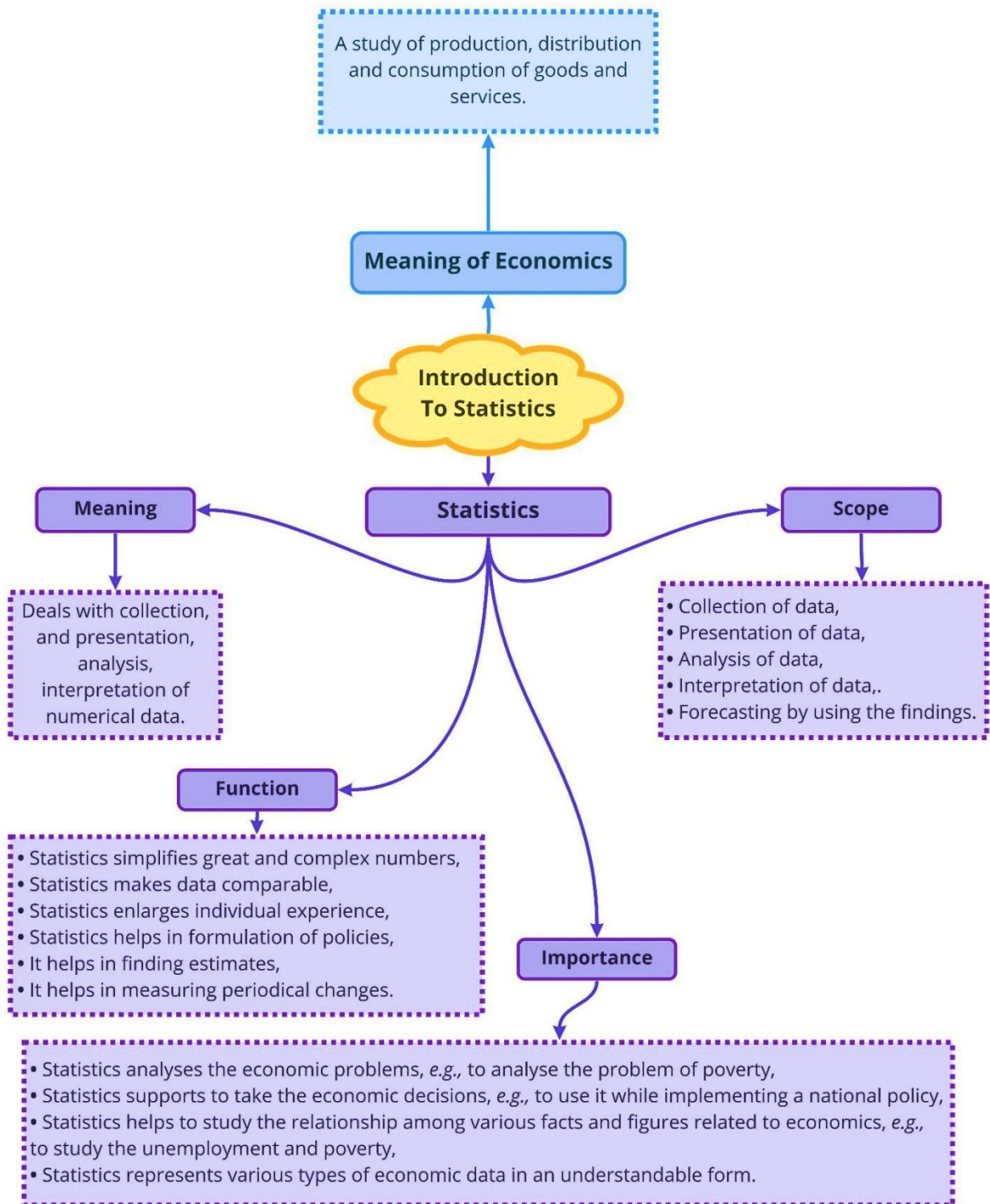
- ❖ Statistics simplifies complexities.
- ❖ It expresses facts in numbers.
- ❖ It presents data in condensed form.
- ❖ Statistics compares different phenomena and reassures relationship between them.
- ❖ Statistics is helpful information of policies.
- ❖ Statistics is helpful in economic forecasting.
- ❖ It facilitates comparisons.
- ❖ It is useful in testing the laws of other sciences.
- ❖ It helps in establishment of correlation between two facts.

Limitations of Statistics

- ❖ Statistics does not study individuals.
- ❖ Statistics results might lead to fallacious conclusions.
- ❖ Statistics deals with quantitative facts only.
- ❖ Statistics laws are true only on averages.
- ❖ Only experts can make the best possible use of statistics.
- ❖ Uniformity and homogeneity of data is essential.
- ❖ Misuse of statistics is indeed its greatest limitation because misuse of statistics is possible.
- ❖ Study of aggregates only
- ❖ Without reference, results may prove to be wrong.



Class : 11th Economics (Statistics)
Chapter-1 : Introduction To Statistics



Important Questions

Multiple Choice questions-

- Arthashashtra originated in:
 - 17th century
 - 18th century
 - 19th century
 - None of the above.
- "Economics is the study of economic welfare" who said this:
 - Marshall
 - Prof. Pigou
 - J. K. Mehta
 - Keynes.
- He is known as father of statistics:
 - Bowley
 - Bodington
 - Gottfried Achenwall
 - None of these.
- Statistics is:
 - Facts
 - Presentation
 - Numerical data
 - None of these.
- Name of the book by Kautilya:
 - Economics
 - Varta
 - Krishna, Valmiki and Vashista
 - None of these.

Fill in the blanks:

- Adam Smith is known as of Economics.
- There is difference between wants and
- Collection is an example of
- Statistics is an art as well as
- When six economists gather they have opinion.

State true or false:

- There are limited wants of humans.
- "Principles of Economics" is written by Pigou.
- Statistical data is numerical data.

- Statistics is not important in speculation market.
- Rules of economics are universal.

Match the columns:

Column A	Column B
1. Optional use of limited resources	(a) Pigou
2. Money is a measuring source	(b) Bowley
3. Statistics is the science of counting	(c) Robbins
4. Collection of data	(d) Samuelson
5. Development related definition of economics	(e) Primary and secondary data.

Very Short Questions:

- According to Marshall, Economics is what type of science?
- Who wrote the book, "Discovery of Nation's Wealth and Reasons"?
- What is statistics in singular?
- Who gave analytical definition of economics?
- From which language statistics has been derived?
- What are economic goods?
- Who is propagator of modern economics?
- Who is known as service provider?
- Who is known as service man?
- What is economic activity?

Short Questions:

- Define statistics.
- What are the stages of statistical study?
- What are the tools used, related to statistical study?
- What are the scopes of statistics?
- Define statistics as a singular noun.
- Define statistics as a plural noun.
- What is inferential statistics?
- What are the two components of the subject matter in statistics?
- What are the three components of economics?
- What is descriptive statistics?



Long Questions:

1. What is Economic Activity?
2. What is statistics?
3. What are Statistical Tools?
4. Subject Matter of Statistics?
5. Limitations of Statistics?
6. What Causes Distrust?

ANSWER KEY

Multiple Choice Answers-

1. (b) 18th century
2. (a) Marshall
3. (b) Bodington
4. (b) Presentation
5. (a) Economics

Fill in the blanks:

1. Father
2. Intensity
3. Statistics
4. Science
5. 7

State true or false:

1. False
2. False
3. True
4. False
5. False

Match the columns:

Column A	Column B
1. Optional use of limited resources	(c) Robbins
2. Money is a measuring source	(a) Pigou
3. Statistics is the science of counting	(b) Bowley
4. Collection of data	(e) Primary and secondary data.
5. Development related definition of economics	(d) Samuelson

Very Short Answers:

1. Social
2. Adam Smith

3. Rules
4. Robbins
5. English
6. The goods which are made by humans are called economic goods.
7. The propagator of modern economics is Prof. Marshall.
8. The person who works for others for remuneration is called service provider.
9. The person who keeps others to do his work for remuneration is called service man.
10. Activities which are performed to earn wealth are called economic activities.

Short Answers:

1. Statistics can be defined as the collection, presentation, classification, analysis, and interpretation of quantitative data.
2. The stages of a statistical study are:
 - Collection of data
 - Organisation of data
 - Presentation of data
 - Analysis of data
 - Interpretation of data
3. The tools used, related to statistical study are:
 - Census or sample technique
 - Tally bar and assembling of data
 - Graphs, tables, and diagrams
 - Average, percentages, regression coefficient, and correlation
 - Average and the degree of relation, percentage, and relation between degree variables
4. The scopes of statistics include:
 - Nature of statistics
 - Subject matter of statistics
 - Limitation of statistics

5. In the singular sense, statistics means the science of statistics or statistical methods. It refers to the techniques or methods relating to the analysis, collection, presentation, classification, and interpretation of quantitative data.
6. In the plural sense, statistics is defined as the information in terms of numerical data or numbers such as employment statistics, statistics concerning public expenditure, population statistics, etc.
7. Inferential statistics refers to the methods by which conclusions are drawn relating to the universe based on a given sample.
8. The two components of the subject matter in statistics are:
 - Descriptive statistics
 - Inferential statistics
9. The three components of economics are consumption, production, and distribution.
10. Descriptive statistics refers to those methods which are used for the collection, presentation as well as analysis of data. These methods relate to such estimations as a measurement of central tendencies, measurement of dispersion, measurement of correlation, etc.

Long Answers:

1. It is an activity which is related to the use of scarce means (also called scarce resources).

Means are always scarce in relation to our wants. Imagine yourself as the richest person on the earth. Still you can't have everything you wish to have at a point of time. It implies the scarcity of your means/ resources in relation to your wants.

Engaged in diverse economic activities, we are performing 'ordinary business of life', according to Alfred Marshall, a great pro founder of Modern Economics. Thus, he defines economics as "the study of mankind in the ordinary business of life."

Scarcity is the Undercurrent of Economic Problem and therefore of Economics

Resources are always scarce in relation to our wants. Also, resources have alternative uses: A ten-rupee note in your pocket may be spent on a cup of coffee or a cold drink. Likewise, a worker may render his services in factory A, rather than B and C. Because, resources are scarce and have alternative uses, we cannot escape from the

problem of allocation of limited means to alternative uses. This is what we call economic problem or the problem of choice.

2. Even to a layman this should not be a difficult question. If asked to define Statistics, we can expect a layman to say that Statistics is something like a store of quantitative information. Yes, it is true. Statistics means quantitative information or quantification of the facts and findings. But, how do we get quantitative information? There must be a system, method or technique to collect quantitative information. Also, statistical information may be a raw information. It needs to be classified, tabulated or it needs to be systematically presented. One must learn the system of presentation and classification of data. Also, there must be a set of methods and techniques to condense the data. May be, we find averages or percentages. And above all, there must be a set of methods or techniques on the analysis and interpretation of quantitative information. A student of economics has to study all these methods and techniques to understand and master the subject matter of Statistics.

Thus, unlike a layman, a student of economics cannot relax taking Statistics just as a pool of quantitative information. Instead he is also to look into the methods or techniques relating to its collection, classification, presentation, analysis as well as interpretation.

In view of such a vastness of the subject matter, Statistics is defined both in singular sense and plural sense, as under:

Statistics—A Plural Noun

In its plural sense, Statistics refers to information in terms of numbers or numerical data, such as Population Statistics, Employment Statistics, Statistics concerning Public Expenditure, etc. However, any numerical information is not Statistics. Example: Ram gets Rs. 100 per month as pocket allowance is not Statistics (it is neither an aggregate nor an average) whereas average pocket allowance of the students of Class X is Rs. 100 per month, or there are 80 students in Class XI compared to just 8 in Class XII of your school are Statistics.

The following table shows a set of data which is Statistics, and another set which is not Statistics. The figures used are hypothetical.



Data which are not Statistics	Data which are Statistics
(i) A cow has 4 legs.	(i) Average height of the 26-plus male people in India is 6 feet compared to 5 feet in Nepal.
(ii) Ram has 200 rupees in his pocket.	(ii) Birth rate in India is 18 per thousand compared to 8 per thousand in USA.
(iii) A young lady was run over by a speeding truck at 100 km per hour.	(iii) Over the past 10 years, India has won 60 test matches in cricket and lost 50.

3. These refer to the methods or techniques used for the collection, organisation and presentation of data, as well as for the analysis and interpretation of data.

Stages of Statistical Study and the Related Statistical Tools

Stages	Statistical Study	Statistical Tools
Stage I	Collection of Data	Census or Sample Techniques
Stage II	Organisation of Data	Array of Data and Tally Bars
Stage III	Presentation of Data	Tables, Graphs and Diagrams
Stage IV	Analysis of Data	Percentages, Averages, Correlation and Regression Coefficients
Stage V	Interpretation of Data	Magnitude of Percentages, Averages and the Degree of Relationship between different economic variables.

4. Subject Matter of Statistics

Subject matter of statistics includes two components: Descriptive Statistics and Inferential Statistics.

The Concept of Universe or Population

It should be interesting for the students of Class XI to note that the concept of universe or population has a specific meaning in Statistics. It refers to the aggregate of all items or units relating to your statistical study. Example:

Universe or population size is 1,000 if you are studying 1,000 students for your statistical study.

(1) **Descriptive Statistics:** Descriptive Statistics refers to those methods which are used for the collection, presentation as well as analysis of data. These methods relate to such estimations as 'measurement of central tendencies' (average mean, median, mode), 'measurement of dispersion' (mean deviation, standard deviation, etc.), 'measurement of correlation', etc. Example: Descriptive statistics is used when you estimate average height of the secondary students in your school. Likewise, descriptive statistics is used when you find that marks in science and mathematics of the students in all classes are intimately related to each other.

(2) **Inferential Statistics:** Inferential Statistics refers to all such methods by which conclusions are drawn relating to the universe or population on the basis of a given sample. (In Statistics, the term universe or population refers to the aggregate of all items or units relating to any subject.) For example, if your class teacher estimates average weight of the entire class (called universe or population) on the basis of average weight of only a sample of students of the class, he is using inferential statistics.

5. In modern times, Statistics has emerged to be of crucial significance in all walks of life. However, it has certain limitations. Thus, writes News home that, "Statistics must be regarded as an instrument of research of great value but barring severe limitations which are not possible to overcome." Following are some notable limitations of Statistics:

(1) **Study of Numerical Facts only:** Statistics studies only such facts as can be expressed in numerical terms, it does not study qualitative phenomena like honesty, friendship, wisdom, health, patriotism, justice, etc.

(2) **Study of Aggregates only:** Statistics studies only the aggregates of quantitative facts. It does not study statistical facts relating to any particular unit. Example: It may be a

statistical fact that your class teacher earns Rs. 50,000 per month. But, as this fact relates to an individual, it is not to be deemed as a subject matter of Statistics. However, it becomes a subject matter of Statistics if we study income of school teachers across all parts of the country, for purpose of finding regional differences in income.

(3) Homogeneity of Data, an essential Requirement:

To compare data, it is essential that statistics are uniform in quality. Data of diverse qualities and kinds cannot be compared. For example, production of food grains cannot be compared with the production of cloth. It is because cloth is measured in meters and food grains in tonnes. Nevertheless, it is possible to compare their value instead of the volume.

(4) Results are True only on an Average:

Most statistical findings are true only as averages. They express only the broad tendencies. Unlike the laws of natural sciences, statistical observations are not error-free. They are not always valid under all conditions. For instance, if it is said that per capita income in India is Rs. 50,000 per annum, it does not mean that the income of each and every Indian is Rs. 50,000 per annum. Some may have more and some may have less,

(5) Without Reference, Results may Prove to be Wrong:

In order to understand the conclusions precisely, it is necessary that the circumstances and conditions under which these conclusions have been drawn are also studied. Otherwise, they may prove to be wrong.

(6) Can be used only by the Experts:

Statistics can be used only by those persons who have special knowledge of statistical methods. Those who are ignorant about these methods cannot make sensible use of statistics. It can, therefore, be said that data in the hands of an unqualified person is like a medicine in the hands of a quack who may abuse it, leading to disastrous consequences. In the words of Yule and Kendall, "Statistical methods are most dangerous tools in the hands of an expert."

(7) Prone to Misuse: Misuse of Statistics is very common. Statistics may be used to support a pre-drawn conclusion even when it is

absolutely false. It is usually said, "Statistics are like clay by which you can make a god or a devil, as you please." Misuse of statistics is indeed its greatest limitation.

6. Distrust of Statistics arises not because there is anything wrong with Statistics as a subject matter. It arises because the users of Statistics tend to manipulate it to suit or support their pre-drawn conclusions or observations. Main causes for the distrust of Statistics are as under:

- (i) Different kinds of Statistics are obtained in respect of a given problem.
- (ii) Statistics can be altered to match the predetermined conclusions.
- (iii) Authentic Statistics can also be presented in such a manner as to confuse the reader.
- (iv) When Statistics are collected in a partial manner, the results are generally wrong. Consequently, people lose faith in them.

However, it may be noted that if Statistics are presented wrongly, then the fault does not lie with Statistics as a subject matter. The fault lies with those people who collect wrong Statistics or those who draw wrong conclusions. Statistics, as such, do not prove anything. They are simply tools in the hands of the statisticians. If a statistician misuses the data, then the blame lies squarely on him and not on the subject matter. A competent doctor can cure a disease by making good use of the medicine but the same medicine in the hands of an incompetent doctor becomes a poison. The fault in this case is not of the medicine but of the unqualified doctor. In the same way, Statistics is never faulty but the fault lies with the users.

In fact, Statistics should not be relied upon blindly nor distrusted outright. "Statistics should not be used as a blind man uses a lamp post for support rather than for illumination, whereas its real purpose is to serve as illumination and not as a support."

In making use of Statistics one should be cautious and vigilant. In the words of King, "The science of Statistics is the most useful servant, but only of great value to those who understand its proper use."

It is the duty of the students of economics to make use of know-how of Statistics to discover the truth rather than to cover the truth.





2

Collection of Data

Definition

Data collection is defined as the procedure of collecting, measuring and analyzing accurate insights for research using standard validated techniques. A researcher can evaluate their hypothesis on the basis of collected data. In most cases, data collection is the primary and most important step for research, irrespective of the field of research. The approach of data collection is different for different fields of study, depending on the required information.

Data is a tool which helps in reaching a sound conclusion by providing information therefore. For statistical investigation, collection of data is the first and foremost.

❖ Sources of Data:

- Primary Source
- Secondary Sources
- Published sources
- Un-published sources

❖ **Primary Data:** Data originally collected in the process of investigation are known as primary data. This is original form of data which are collected for the first time. It is collected directly from its source of origin.

Methods of collecting primary data –

There are three basic ways of collecting data:

- Personal interview OR Direct Personal Investigation.
- Mailing (questionnaire surveys).
- Telephone interviews.
- Indirect verbal investigation.
- Information from local sources.
- Enumerator method.

Secondary data It refers to collection of data by some agency, which already collected the data and processed. The data thus collected is called secondary data.

Point of difference between Primary and Secondary data -

- Accuracy,
- Originality,
- Cost,
- Need of modification

BASIS FOR COMPARISON	PRIMARY DATA	SECONDARY DATA
Meaning	Primary data refer to the first hand data gathered by the researcher himself.	Secondary data means data collected by someone else earlier.
Data	Real time data	Past data
Process	Very involved	Quick and easy
Source	Surveys, observations, experiments, questionnaire, personal interview, etc.	Government publications, websites, books, journal articles, internal records etc.
Cost effectiveness	Expensive	Economical
Collection time	Long	Short
Specific	Always specific to the researcher's needs.	May or may not be specific to the researcher's need.
Accuracy and Reliability	More	Relatively less

Sources of secondary data -

Secondary Data: Using existing data generated by large government Institutions, healthcare facilities etc. as part of organizational record keeping. The data is then extracted from more varied datafiles.

Published sources:

- Govt. publication.
- semi-Govt. Publication.
- Reports of committees & commissions.
- Private publications e.g., Journals and News papers research institute, publication of trade association.
- International publications.

Unpublished Sources -

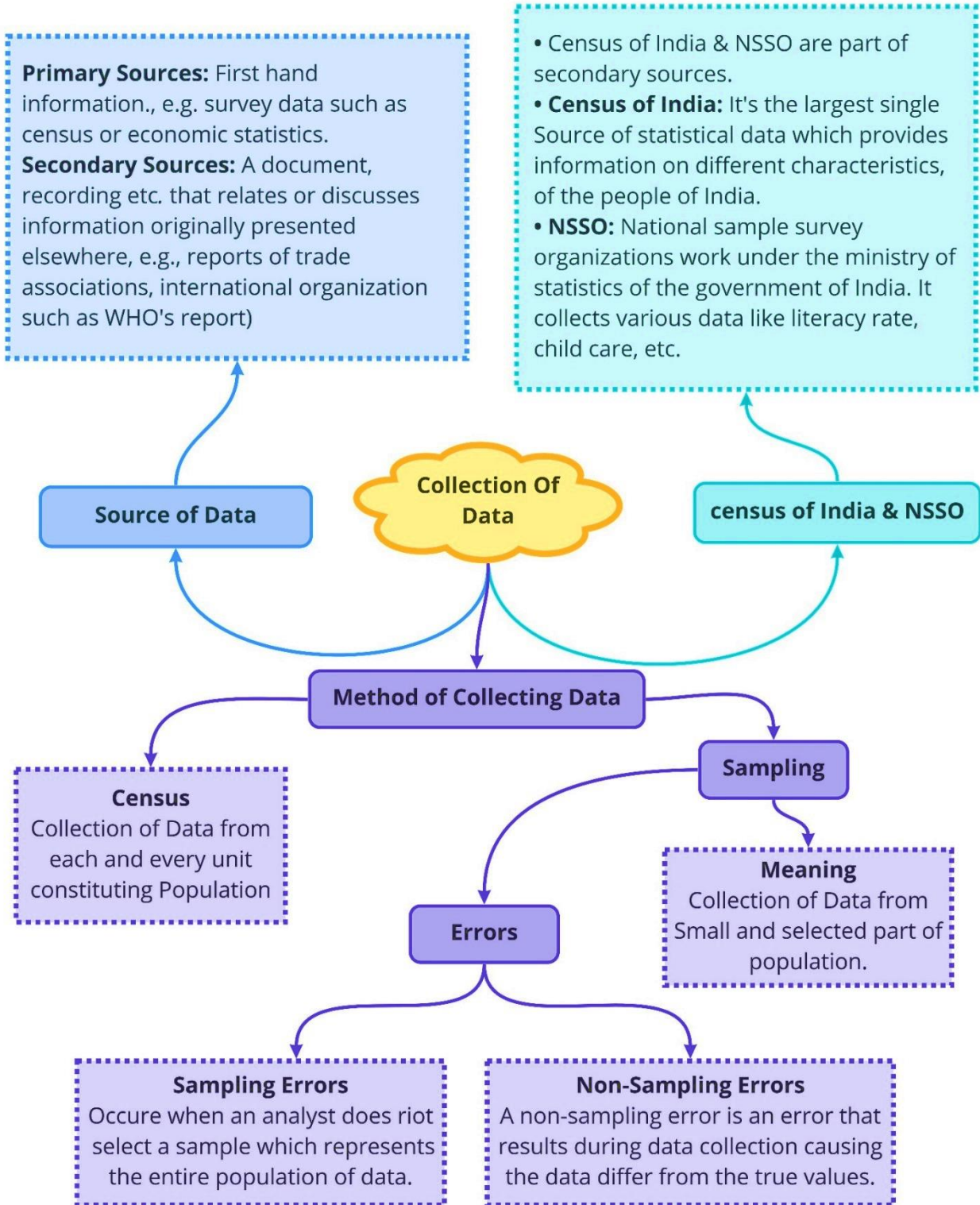
The statistical data needn't always be published. There are various sources of unpublished statistical material such as the records maintained by private firms, business enterprises, scholars, research workers, etc. They may not like to release their data to any outside agency.

Other source : web-site

Pilot Survey: Before sending the questionnaire to the information. It should be pretested. As a result of its short comings if any, can be removed. Such pretesting named as pilot survey



Class : 11th Economics (Statistics)
Chapter-2 : Collection Of Data



Important Questions

Multiple Choice Questions-

- Stratified sample is preferred where:
 - Population is perfectly homogeneous
 - Population is non-homogeneous
 - Random sampling is not possible
 - Small samples are required
- Data collected for the first time from the source of origin is called:
 - primary data
 - secondary data
 - internal data
 - none of these
- Census method is suitable for that investigation in which:
 - the size of population is large
 - high degree of accuracy is not required
 - there are widely diverse items
 - intensive examination of diverse items is not required
- Which of the following factor(s) are considered when comparison between sampling and census method is made?
 - Area of survey
 - Accuracy of data
 - Cost of collection
 - All of these
- Reliability of sampling data depends on:
 - size of sample
 - method of sampling
 - training of enumerators
 - all of these
- Under random sampling, each item of the universe has __ chance of being selected.
 - equal
 - unequal
 - zero
 - none of these
- Which of the following methods is used for the estimation of population in a country?
 - Census method
 - Sampling method
 - Both (a) and (b)
 - None of these
- Personal bias is possible under:
 - random sampling
 - purposive sampling
 - stratified sampling
 - quota sampling
- If the investigator wants to select a sample on the basis of diverse characteristics of the population, which method should he use?
 - Convenience sampling method
 - Quota sampling method
 - Stratified sampling method
 - Both (b) and (c)
- For drawing lottery _____ sampling is used.
 - random
 - purposive
 - stratified
 - quota
- Which of the following methods is used for the estimation of population in country`
 - Sampling Method
 - Census Method
 - Both (a) and (b)
 - Neither (a) nor (b)
- What kind of data are contained in the census of population and national income estimates, for the government?
 - Primary data
 - Secondary data
 - Internal data
 - None of these
- The data collected on the height of a group of students after recording their heights with a measuring tape are:
 - Primary data
 - Continuous data
 - Discrete data
 - Secondary data
- Which of the following is a method of secondary data collection?
 - Direct personal investigation
 - Direct oral investigation
 - Collection of information through questionnaire
 - None of these



15. In random sampling:
- Each element has equal chance of being selected
 - Sample is always full of bias
 - Cost involved is very less
 - Cost involved is high

Very Short Questions:

- Define primary data.
- Define secondary data.
- What are the two sources of data?
- Mention two sources of secondary data.
- In what parameters is the statistical information published in the census of India?
- Mention two demerits of indirect oral investigation.
- The progress report of a railway published by the railway department is what kind of data?
- When is a direct personal investigation suitable for primary data collection?
- When are the qualities of a good Questionnaire?
- Why is a pilot survey important?
- What is the universe in statistics?
- Define sample.

- Define the census method.
- Explain the sample method.
- What do you mean by random sampling?
- What is purposive or deliberate sampling?
- Define stratified and mixed sampling?
- Explain systematic sampling.
- What is quota sampling?
- What is convenience sampling?

Short & Long Questions:

- Primary Source?
- Secondary Source?
- Principal Differences between Primary and Secondary Data?
- Direct Personal Investigation?
- Indirect Oral Investigation
- Difference between Direct Personal Investigation and Indirect Oral Investigation
- Information from Local Sources or Correspondents
- Collection of secondary data?
- Statistical Errors: Sampling and Non-Sampling Errors

ANSWER KEY

Multiple Choice Answers-

- B
- A
- C
- D
- D
- A
- A
- B
- D
- A
- B
- B
- A
- D
- A

Very Short Answers:

- Primary data is the collection of data collected by the investigator for his own purpose for the first time. These are collected from the source of origin.
- According to Wessel, "Data collected by another person is known as secondary data". It is known as secondary data as it has already been collected by somebody else. These data are accessible in the form of a published and unpublished report.
- The two sources of data are:
 - Primary source
 - Secondary source
- The two sources of secondary data are:
 - Government publication
 - Semi-government publication
- The statistical information is published in the following parameters in the census of India

- Population projection
 - Sex composition of a population
 - Density of population
 - Size, growth rate, and distribution of people in India
6. The two demerits of indirect oral investigation are:
- Less accurate
 - Biased
 - Doubtful conclusion
7. The progress report of a railway published by the railway department is secondary data.
8. The direct personal investigation method is suitable for collecting primary data only on the following situations:
- When the investigation is confined and less
 - When an authentic and accurate information is required
 - When the data is to be kept secret
 - When the direct contact with information is needed
9. A good Questionnaire should have the following qualities:
- Less number of Questions
 - Should be clear
 - Proper order of Question
 - Non-controversial
 - Questions related to the topic
 - Request for return
10. A pilot survey is essential because of the following:
- It helps in assessing the quality and suitability of Questions.
 - It evaluates the performance of enumerators.
 - It helps in designing a set of rules for the investigator.
 - It estimates the time and cost involved in the final survey.
11. In statistics, the term universe or population indicates an aggregate of items studied for investigators.
12. Sample is a collection of an item from the population that represents the characteristics of the population.
13. It is a method of collecting data where each item related to the problem of the investigation is collected.
14. It is a process of collecting data in which the sample of a group of items are examined, and conclusions are drawn on their basis.
15. In this method, every item of the universe has an equal chance of being selected in the sample.
16. It is a sampling method where the investor chooses the sampling items according to his opinion, and it is the best for the population.
17. In this method, the universe is divided into two groups having different characteristics, and the items are selected for each group, hence the entire group is represented.
18. In systematic sampling, population units are arranged according to the alphabets, numbers, and geography. Here, every n th numerical item is selected as a sample.
19. Here, the universe is divided into two sections or groups in terms of their characteristics.
20. In this method, sampling is done according to the investigator's convenience.

Short & Long Answers:

1. You want to know about the quality of life of the people in your town. You may like to ascertain the quality of life in terms of per capita expenditure of different households in your town. You decide to collect the basic data yourself through statistical survey(s), of course with the help of investigators or field workers. While doing this exercise you are relying on primary source of the data. Thus, primary source of data implies collection of data from its source of origin. It offers you firsthand quantitative information relating to your statistical study. You or your team of investigators are contacting the respondents (people offering basic information) and obtaining the desired quantitative information on per capita expenditure of different households in your town. Primary source of data implies collection of data from its source of origin. It offers you first-hand quantitative information relating to your statistical study.
2. Secondary Source of collection of data implies obtaining the relevant statistical information from an agency, or an institution which is already in possession of that information. To continue with the previous example, data relating to the quality of life of the people of your town (or the data on per capita expenditure) may have already been collected by the State Government. You can simply approach the concerned Government department



and request for the desired information. This will be a Secondary Source of data for you. Thus, secondary source implies that the desired statistical information already exists and you are simply to collect it from the concerned agency or the department. You are not to conduct statistical survey(s) yourself and you are not to contact the respondents (people offering basic information). OT course, you are not getting first hand information relating to your statistical study. You are simply relying on the information which is already existing.

Secondary source of data implies collection of data from some agency or institution which already happens to have collected the data through statistical survey(s). It does not offer you first-hand information relating to your statistical study. You are to rely on the information which is already existing.

3. The following are some principal differences between primary and secondary data:

- 1) **Difference in Originality:** Primary data are original because these are collected by the investigator from the source of their origin. Against this, secondary data are already in existence and therefore, are not original.
- 2) **Difference in Objective:** Primary data are always related to a specific objective of the investigator. These data, therefore, do not need any adjustment for the concerned study. On the other hand, secondary data have already been collected for some other purpose. Therefore, these data need to be adjusted to suit the objective of study in hand.
- 3) **Difference in Cost of Collection:** Primary data are costlier in terms of time, money and efforts involved than the secondary data. This is because primary data are collected for the first time from their source of origin. Secondary data are simply collected from the published or unpublished reports. Accordingly, these are much less expensive.

Of course, it may be noted that, there are no fundamental differences between primary data and secondary data. Data are data, whether primary or secondary. These are classified as primary or secondary just on the basis of their collection: first-hand or second-hand. Thus, a particular set of data when collected by the investigator for a specific purpose from the source of origin, would be

primary data. And the same set of data, when used by some other investigator for his own purpose, would be known as secondary data. Thus, Secrist has rightly pointed out, "The distinction between primary and secondary data is one of the degree. Data which are primary in the hands of one party may be secondary in the hands of other."

Primary and Secondary Data—The Basic Difference

- If we are collecting data from its source of origin, for the first time, it is primary data.
- If we are using data which have already been collected by somebody else, it is secondary data.

Note: If you are getting data from somebody else who collected it from its source of origin but did not use it for his own study, it will be deemed as primary data.

4. The direct personal investigation is the method by which data are personally collected by the investigator from the informants. In other words, the investigator establishes direct relation with the persons from whom the information is to be obtained. The success of this method, however, requires that the investigator should be very diligent, efficient, impartial and tolerant.

Direct contact with the workers of an industry to obtain information about their economic conditions is an example of this method.

Suitability

This method of collecting primary data is suitable particularly when:

- (i) the field of investigation is limited or not very large.
- (ii) a greater degree of originality of the data is required.
- (iii) information is to be kept secret.
- (iv) accuracy of data is of great significance, and
- (v) when direct contact with the informants is required.

Merits

Data, thus, collected have the following merits:

- (i) **Originality:** Data have a high degree of originality.
- (ii) **Accuracy:** Data are fairly accurate when personally collected.

- (iii) **Reliability:** Because the information is collected by the investigator himself, reliability of the data is not doubted.
- (iv) **Related Information:** When in direct contact with the informants, the investigator may obtain other related information as well.
- (v) **Uniformity:** There is a fair degree of uniformity in the data collected by the investigator himself from the informants. It facilitates comparison.
- (vi) **Elastic:** This method is fairly elastic because the investigator can always make necessary adjustments in his set of questions.

Demerits

However, the method of direct personal investigation suffers from certain demerits, as under:

- (i) **Difficult to Cover Wide Areas:** Direct personal investigation becomes very difficult when the area of the study is very wide.
- (ii) **Personal Bias:** This method is highly prone to personal bias of the investigator. As a result, the data may lose their credibility.
- (iii) **Costly:** This method is very expensive in terms of the time, money and efforts involved.
- (iv) **Limited Coverage:** In this method, area of investigation is generally small. The results are, therefore, less representative. This may lead to wrong conclusions.

5. Indirect oral investigation is the method by which information is obtained not from the persons regarding whom the information is needed. It is collected orally from other persons who are expected to possess the necessary information, these other persons are known as witnesses. For example, by this method, the data on the economic conditions of the workers may be collected from their employers rather than the workers themselves.

Suitability

This method is suitable particularly when:

- (i) the field of investigation is relatively large.
- (ii) it is not possible to have direct contact with the concerned informants.
- (iii) the concerned informants are not capable of giving information because of their ignorance or illiteracy.

- (iv) investigation is so complex in nature that only experts can give information.

This method is mostly used by government or non-government committees or commissions.

Merits

Some of the notable merits of this method are as under:

- (i) **Wide Coverage:** This method can be applied even when the field of investigation is very wide.
- (ii) **Less Expensive:** This is relatively a less expensive method as compared to Direct Personal Investigation.
- (iii) **Expert Opinion:** Using this method an investigator can seek opinion of the experts and thereby can make his information more reliable.
- (iv) **Free from Bias:** This method is relatively free from the personal bias of the investigator.
- (v) **Simple:** This is relatively a simple approach of data collection.

Demerits:

However, there are some demerits, as under:

- (i) **Less Accurate:** The data collected by this method are relatively less accurate. This is because the information is obtained from persons other than the concerned informants.
- (ii) **Biased:** There is possibility of personal bias of the witnesses giving information.
- (iii) **Doubtful Conclusions:** This method may lead to doubtful conclusions due to carelessness of the witnesses.

6. The difference between direct personal investigation and indirect oral investigation is as under:

- i. In the case of direct personal investigation, the investigator establishes direct contact with the informants. On the other hand, in the case of indirect oral investigation, information is obtained by contacting other than those about whom information is sought.
- ii. Direct Personal Investigation is generally possible when the field of investigation is small. On the other hand, indirect oral investigation is generally preferred when the field of investigation is relatively large.



- iii. In the Direct Personal Investigation, the investigator must be well versed in the language and cultural habits of the informants. There is no such requirement in the case of Indirect Oral Investigation.
 - iv. Direct investigation is relatively costlier than the indirect investigation.
7. Under this method, the investigator appoints local persons or correspondents at different places. They collect information in their own way and furnish the same to the investigator.

Suitability

This method is suitable particularly when:

- (i) regular and continuous information is needed.
- (ii) the area of investigation is large.
- (iii) the information is to be used by journals, magazines, radio, TV, etc. and
- (iv) a very high degree of accuracy of information is not required.

Merits

Principal merits of this method are as under:

- (i) **Economical:** This method is quite economical in terms of time, money or efforts involved.
- (ii) **Wide Coverage:** This method allows a fairly wide coverage of investigation.
- (iii) **Continuity:** The correspondents keep on supplying almost regular information.
- (iv) **Suitable for Special Purpose:** This method is particularly suitable for some specialpurpose investigations, e.g., price quotations from the different grain markets for the construction of Index Number of agricultural prices.

Demerits

Following are some notable demerits of this method:

- (i) **Loss of Originality:** Originality of data is sacrificed owing to the lack of personal contact with the respondents.
- (ii) **Lack of Uniformity:** There is lack of uniformity of data. This is because data is collected by a number of correspondents.
- (iii) **Personal Bias:** This method suffers from the personal bias of the correspondents.
- (iv) **Less Accurate:** The data collected by this method are not very accurate.

- (v) **Delay in Collection:** Generally, there is a delay in the collection of information through this method.

8. There are two main sources of secondary data:

(1) Published Sources

Some of the published sources of secondary data are:

- (i) **Government Publications:** Ministries of the Central and State Governments in India publish a variety of Statistics as their routine activity. As these are published by the Government, data are fairly reliable. Some of the notable Government publications on Statistics are: Statistical Abstract of India, Annual Survey of Industries, Agricultural Statistics of India, Report on Currency and Banking, Labour Gazette, Reserve Bank of India Bulletin, etc.
- (ii) **Semi-Government Publications:** Semi-Government bodies (such as Municipalities and Metropolitan Councils) publish data relating to education, health, births and deaths. These data are also fairly reliable and useful.
- (iii) **Reports of Committees and Commissions:** Committees and Commissions appointed by the Government also furnish a lot of statistical information in their reports. Finance Commission, Monopolies Commission, Planning Commission are some of the notable commissions in India which supply detailed statistical information in their reports.
- (iv) **Publications of Trade Associations:** Some of the big trade associations, through their statistical and research divisions, collect and publish data on various aspects of trading activity. For example, Sugar Mills Association publishes information regarding sugar mills in India.
- (v) **Publications of Research Institutions:** Various universities and research institutions publish information as findings of their research activities. In India, for example, Indian Statistical Institute, National Council of Applied Economic Research publish a variety of statistical data as a regular feature.

(vi) **Journals and Papers:** Many newspapers such as 'The Economic Times' as well as magazines such as Commerce, Facts for You also supply a large variety of statistical information.

(vii) **Publications of Research Scholars:** Individual research scholars also sometimes publish their research work containing some useful statistical information.

(viii) **International Publications:** International organisations such as UNO, IMF, World Bank, ILO, and foreign governments etc., also publish a lot of statistical information. These are used as secondary data.

(2) Unpublished Sources

There are some unpublished secondary data as well. These data are collected by the government organisations and others, generally for their self use or office record. These data are not published. This unpublished numerical information may, however, be used as secondary data.

A Note of Caution for the Users of Secondary Data Users of secondary data must check:

- (i) reliability of data,
- (ii) suitability of data, and
- (iii) adequacy of data.

9. Statistical errors are broadly classified as (i) sampling errors, and (ii) non-sampling errors. Following are the details:

(i) **Sampling Errors:** These are related to the size or nature of the sample selected for the study. Due to a very small size of the sample selected for study or due to nonrepresentative nature of the sample, the

estimated value may differ from the actual value of a parameter. The error thus emerging, is called sampling error. For example, if the estimated value of a parameter is found to be 10 while the actual/true value is 20 then, the sampling error = estimated value of the parameter – true value of the parameter = $10 - 20 = -10$.

(ii) **Non-sampling Errors:** These are errors related to the collection of data. These are of the following types:

Error of Measurement: Error of measurement may occur due to- (a) difference in the scale of measurement, and (b) difference in the rounding off procedure adopted by different investigators.

Error of Non-response: This arises when the respondents do not offer the required information. **Error of Misinterpretation:** This arises when the respondent fails to interpret the questions in the questionnaire.

Error of Calculation or Arithmetical Error: It occurs in the course of addition, subtraction or multiplication of data.

Error of Sampling Bias: It occurs when, for some reason or the other, a part of target population, cannot be included in the choice of a sample.

Larger the field of investigation or larger the population size, greater is the possibility of errors related to the collection of data, or data acquisition. It must be noted here that a non-sampling error is more serious than a sampling error. Because a sampling error can be minimised by opting for a larger sample size. No such possibility exists in case of nonsampling errors.



❖ **Continuous Series:** A variable is called continuous if it can take any value in a given range. In constructing continuous series we come across terms like:

- **Class:** Each given interval is called a class e.g., 0-5, 5-10.
- **Class limit:** There are two limits upper limit and lower limit.
- **Class interval:** Difference between upper limit and lower limit.
- **Rang:** Difference between upper limit and lower limit.

$$\frac{\text{Upper limit} - \text{Lower limit}}{2}$$

- **Mid-point or Mid Value:**
- **Frequency:** Number of items [observations] falling within a particular class.

❖ **Exclusive Series:** Excluding the upper limit of these classes, all the items of the class are included in the class itself. E.g.,

❖ **Inclusive:**

Marks	0-10	10-20	20-30	30-40
Number of Students	2	5	2	1

- **Series:** Upper class limits of classes are included in the respective classes. E.g.,

Marks	0-9	10-19	20-29
Name of Students	2	5	2

- **Open End Classes:**

The lower limit of the first class and upper limit of the last class are not given. E.g.,

Marks	Below 20	20-30	30-40	40-50	50 and above
Number of Students	7	6	12	5	3

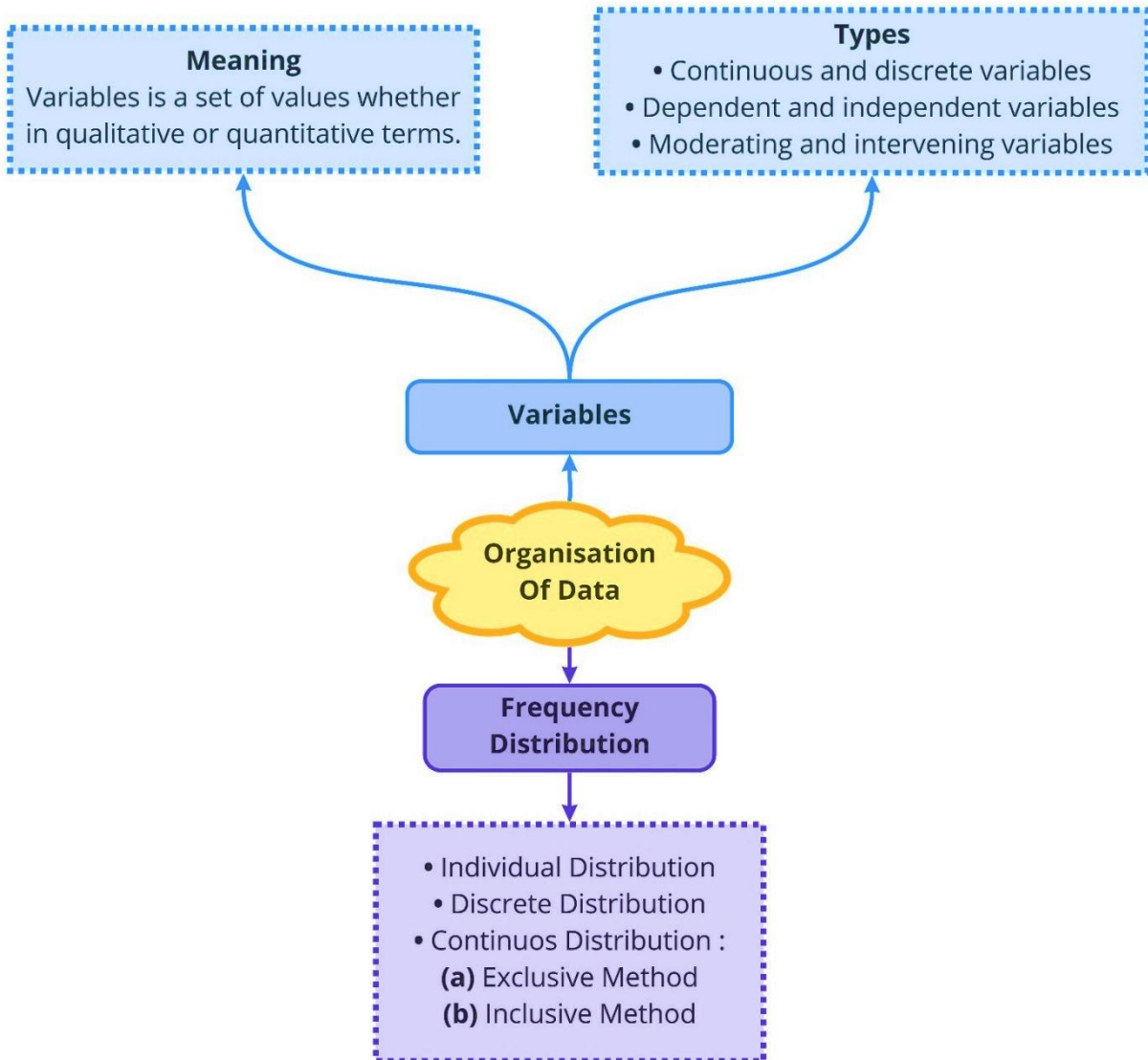
❖ **Cumulative Frequency Series:** It is obtained by successively adding the frequencies of the values of the classes according to a certain law.

- **'Less than' Cumulative Frequency Distribution:** The frequencies of each class-interval are added successively.
- **'More than' Cumulative Frequency Distribution:** The more than cumulative frequency is obtained by finding the cumulative totals of frequencies starting from the highest value of the variable to the lowest value.

Marks	No. of Students	Marks	No. of Students	Marks	No. of Students
0-10	2	Less than 10	2	More than 0	50
10-20	5	Less than 20	7	More than 10	48
20-30	10	Less than 30	17	More than 20	43
30-40	12	Less than 40	29	More than 30	33
40-50	17	Less than 50	46	More than 40	21
50-60	4	Less than 60	50	More than 50	4



Class : 11th Economics (Statistics)
Chapter-3 : Organisation Of Data



Important Questions

Multiple Choice questions-

- The characteristic of a fact that can be measured in the form of numbers is called:
 - frequency
 - variable
 - attribute
 - none of these
- Frequency of a variable is always:
 - A fraction
 - In percentage
 - An integer
 - None of these
- A series showing the sets of all values in classes with their corresponding frequencies is known as:
 - Grouped frequency distribution
 - Cumulative frequency distribution
 - Simple frequency distribution
 - None of the above
- A series in which every class interval excludes items corresponding to its upper limit is called:
 - exclusive series
 - inclusive series
 - both (a) and (b)
 - none of these
- An open-end series is that series in which:
 - lower limit of the first class interval is missing
 - upper limit of the last class interval is missing
 - both (a) and (b)
 - none of these
- According to tally bar method, which of the following symbols indicate the frequency of five?
 - III
 - II
 - III
 - None of these
- The Frequency distribution of a continuous variable is known as:
 - Grouped frequency distribution
 - Simple frequency distribution
 - Either (a) or (b)
 - Both (a) and (b)
- In an ordered series, the data are:
 - In descending order
 - In ascending order
 - Either (a) or (b)
 - None of these
- In a series, the number of times an item occurs is known as:
 - number
 - class frequency
 - frequency
 - cumulative frequency
- The difference between upper limit and lower limit of a class is known as:
 - range
 - magnitude of a class interval
 - frequency
 - class limits
- Which of the following equations is correct?
 - $s = r + n$
 - $s = r - n$
 - $s = r \times n$
 - $s = rn$

[s = Size of the class; r = Range; n = Number of classes]
- Drinking habit of a person is:
 - An attribute
 - A discrete variable
 - A variable
 - A continuous variable
- An attribute is:
 - A qualitative characteristic
 - A measurable characteristic
 - A quantitative characteristic
 - All these
- Nationality of a student is:
 - An attribute
 - A discrete variable
 - A continuous variable
 - Either (a) or (c)



15. Why is it true that classes in frequency distributions are all inclusive`
- No data point falls into more than one class
 - There are always more classes than data point
 - All data fit into one class or another
 - All of these

Very Short Questions:

- What is the classification?
- Define variable
- Define individual series.
- Explain the discrete series.
- What do you mean by frequency distribution or frequency series?
- What is the frequency?
- Define the class limit.
- Explain the magnitude of the class interval.
- What is an exclusive series?
- What is an inclusive series?

Short & Long Questions:

- What is classification?
- What is the Principal Objective of Classification of Data?
- Basis of Classification?
- Concept of variable?
- Raw Data?
- Types of Statistical Series?
- Types of frequency distribution?
- What is the basic difference between Exclusive Series and Inclusive Series?
- What is an Open-End Series?

Assertion Reason Question:

- In these questions, a statement of assertion followed by a statement of reason is given. Choose the correct answer out of the following choices.
 - Both Assertion (A) and Reason (R) are true, and Reason (R) is the correct explanation of Assertion (A)
 - Both Assertion (A) and Reason (R) are true, and Reason (R) is not the correct explanation of Assertion (A)
 - Assertion (A) is true, but Reason (R) is false.
 - Assertion(A) is false, but Reason (R) is true.

Assertion (A): In the case of exclusive class intervals, upper limit is not included.

Reason (R): In the case of exclusive class intervals, we have to decide in advance which class limit is to be excluded.

- In these questions, a statement of assertion followed by a statement of reason is given. Choose the correct answer out of the following choices.
 - Both Assertion (A) and Reason (R) are true, and Reason (R) is the correct explanation of Assertion (A)
 - Both Assertion (A) and Reason (R) are true, and Reason (R) is not the correct explanation of Assertion (A)
 - Assertion (A) is true, but Reason (R) is false.
 - Assertion(A) is false, but Reason (R) is true.

Assertion (A): Classification brings order to raw data.

Reason (R): Classification should be formed in such a way that the class mark of each class comes as close as possible, to a value around which the observations in a class tend to concentrate.

ANSWER KEY

Multiple Choice Answers-

- | | |
|------|-------|
| 1. B | 8. C |
| 2. C | 9. C |
| 3. A | 10. B |
| 4. A | 11. D |
| 5. C | 12. A |
| 6. D | 13. A |
| 7. A | 14. A |
| | 15. C |

Very Short Answers:

1. Classification means a division of classes on the basis of their diversity and similarity.
2. Variable can be defined as a situation where its value keeps changing and is capable of being measured.
3. It is a series where items are singly listed. In this series, there is no class and items.
4. In this series, data are presented a such a way that it shows the exact measurement of items.
5. It is a series that cannot have an exact measure. The items assume a range of value and are placed within the limit or range.
6. Frequency is the number of times an item occurred or is repeated.
7. The extreme value is the class limit. Every class has two limits, lower and upper limit.
8. Magnitude of the class interval is the upper and lower limit of a class
9. It is a series where every class interval removes items related to an upper limit
10. An inclusive series is a series that includes all the items until the upper limit.

Short & Long Answers:

1. There are 200 families in your locality. You have collected data regarding their income, expenditure, education, religion, size of family, etc. But this data will be of little use unless you know how many families are educated and how many are uneducated. How many families earn an income exceeding Rs. 5,000 per month and how many earn Rs. 500 or less per month.

In other words, in order to make the raw data meaningful, these must be classified on the basis of their different characteristics, such as educated families and uneducated families, rich families and poor families, etc.

2. **It is to capture and distinctively present the diverse characteristics of data.**

- (i) Data are divided into different groups. For example, on the basis of education, persons may be classified as educated and uneducated.
- (ii) Data are grouped or classified on the basis of their class similarities. All similar units are put in one class and as the similarity changes, class also changes. Objectives of Classification

Main objectives of Classification are as under:

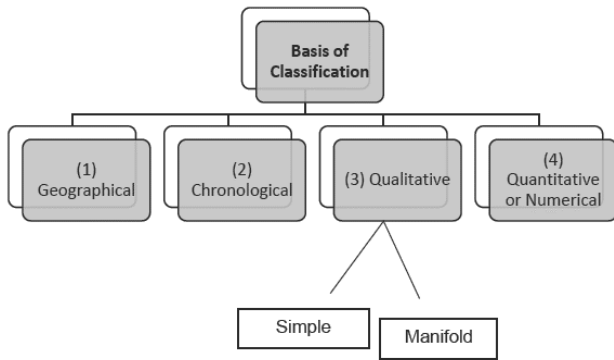
- (1) **Brief and Simple:** Main objective of classification is to present data in a form that appears to be brief and simple.
- (2) **Utility:** Classification enhances utility of the data as it brings out similarity within the diverse set. of data.
- (3) **Distinctiveness:** Classification renders obvious differences among the data more distinctly.
- (4) **Comparability:** It makes data comparable and estimative.
- (5) **Scientific Arrangement:** Classification facilitates arrangement of data in a scientific manner which increases their reliability.
- (6) **Attractive and Effective:** Classification makes data more attractive and effective.

Characteristics of a Good Classification

- (1) **Comprehensiveness:** Classification of the raw data should be so comprehensive that each and every item of the data gets into some group or class. No item should be left out.
- (2) **Clarity:** Classification of the raw data into classes should be absolutely clear and simple. That is, there should be no confusion about the placement of any item in a group.
- (3) **Homogeneity:** All items in a group or class must be homogeneous or similar to each other.
- (4) **Suitability:** The composition of the classes must suit the objective of enquiry. For example, in order to determine the income and expenditure of the students in a school, their classification on the basis of weight or marital status would make no sense. The data must be classified on the basis of different levels of income and expenditure.
- (5) **Stability:** A particular kind of investigation should be based on the same set of classification. This base should not change with each investigation.
- (6) **Elastic:** Classification should be elastic. There should be a scope for change in the classification, depending on the change in purpose or objective of the study.

3. Basis of Classification

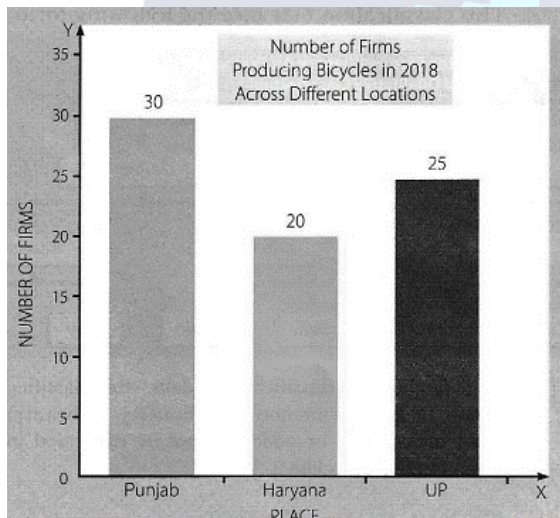
There may be different basis of classifying a statistical information as shown in chart below.



(1) Geographical (or Spatial) Classification: This classification of data is based on the geographical or locational differences of the data. To illustrate, data relating to the number of firms producing bicycles in India would be classified as under:

Table 1. Number of Firms Producing Bicycles in 2018 across Different Locations

Place	Number of Firms
Punjab	30
Haryana	20
UP	25



(2) Chronological Classification: When data are classified on the basis of time, it is known as chronological classification. This is illustrated in the following table 2.

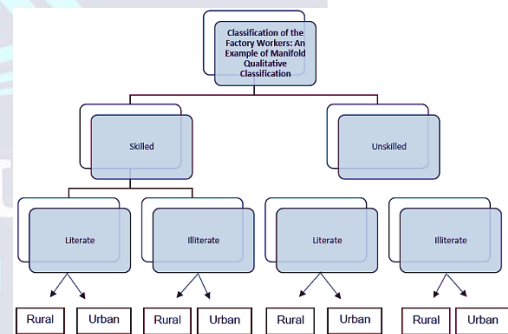
Table 2 Sales of a Firm (2016-2018)

Year	Sales (Rs.)
2016	80 Lakh
2017	90 Lakh
2018	95 Lakh

(3) Qualitative Classification: This classification is according to Qualities or Attributes of the data. For example, data may be classified on the basis of occupation, religion, level of intelligence of the population. This classification may be of two types:

(i) Simple Classification: It is called classification according to dichotomy. This is because data are divided on the basis of existence or absence of a quality. Male-female, healthy-unhealthy, educated-uneducated, are examples of dichotomy.

(ii) Manifold Classification: When classification according to quality of data involves more than one characteristic, it is called manifold classification or multiple classification. As a result of it, there may be more than two classes. To illustrate, factory workers may be classified as 'skilled' and 'unskilled'. These may be further classified as literate or illiterate and still further as rural or urban. This classification may take the following form:



[Note: In qualitative classification, data are classified on the basis of a phenomenon (like honesty or beauty) which is not measurable or which cannot be expressed in terms of quantitative units like 2, 3 or 4.]

(4) Quantitative or Numerical Classification: Classification is done on the basis of numerical values of the facts. A number of classes are framed keeping in view the lowest and highest value as well as the range of values in the data. Each class of a set of data refers to a phenomenon like 'wages' or 'profits' in the automobile industry which can be expressed in figures like Indian rupees. Table 3 below is an illustration of quantitative classification:

Table 3. Annual Profit of Small-Scale Firms in the State of UP: Hypothetical Data, just for an illustration of Quantitative Classification

Annual Profit (Rs.)	Number of Firms
0 - 1,00,000	5
1,00,000 – 2,00,000	150
2,00,000 – 3,00,000	1500
3,00,000 – 4,00,000	800
4,00,000 – 5,00,000	400
Above 5,00,000	200

In the above classification, profit is the phenomenon under study. It is a quantifiable phenomenon. Hence, it is called quantitative classification of data.

It is important to note that the phenomenon under study (like profit in the above illustration) assumes different values over time or across different regions. When a phenomenon assumes different values it is called 'variable' in Statistics. Accordingly, quantitative classification is also called 'classification by variables.'

4. **CONCEPT OF VARIABLE** - A characteristic or a phenomenon which is capable of being measured and changes its value overtime is called a variable. Thus, a variable refers to that quantity which is subject to change and which can be measured by some unit. If we measure the weight of students of Class XI, then the weight of the students will be called variable. A variable may be either discrete or continuous.

Principal difference between Discrete Variable and Continuous Variable

It is that while discrete variable assumes values in complete numbers like 2, 4, 5 and 8, continuous variables assume values in fractions like 2.4, 4.6 and 6.8 or values in some range like 2-4, 4-6 etc.

- (1) **Discrete Variable:** Discrete variables are those variables that increase in jumps or in complete numbers. For example, the number of students in Class XI could be 1, 2, 3, 10, 11, 15 or 20 etc. but cannot be 1 1 4 , 1 1 2 , 1 3 4 etc. In other words, discrete variables are expressed in terms of complete numbers, or one may simply say that values of these variables are in complete numbers as 1,2, 3 and not continuous as between 1 to 1 1 4 or a 1 3 4 and so on.

- (2) **Continuous Variable:** Variables that assume a range of values or increase not in jumps but continuously or in fractions are called continuous variables. For example, height of the boys in a school is expressed as 5'1", 5'2", 5'3", and so on. In short, while the values of discrete variables are in complete numbers (f 2, 3, etc.), values of continuous variables are in fractions (5'4", 5'2", etc.) or are in any range such as 10-15, 15-20, etc.

Difference between Variable and Attribute

Ordinarily, anything that varies/changes over time is taken as a variable. But not in Statistics. Colour of your hair may change over time. Is it a variable? No, not at all. Why? Because this change cannot be numerically expressed. In Statistics, only that change of an object is taken as a variable which can be numerically expressed For example, average height of the students of Class X in the year 2005 is found to be (say) 5'6" compared to 55" in the previous year. Qualitative change, like a change in IQ level of the students of Class X is called attribute'. A change in the attributes is only qualitatively expressed as good, excellent and outstanding. Qualitative changes can at best be ranked as 1, 2, 3 where 1 stands for outstanding, 2 stands for excellent and 3 stands for good.

5. A mass of data in its crude form is called raw data. It is an unorganised mass of the various items. These are yet to be organised by the investigator. To illustrate, marks obtained by 30 students of class XI in Statistics, may be expressed as in Table 4.

30	20	40	20	15	20
25	10	20	15	25	20
15	45	10	30	20	25
30	20	30	20	15	35
25	10	25	15	35	10

Data presented in this table are raw data. These are not homogeneous data or the data classified into different groups or classes with similarities. No meaningful conclusion is possible from this data. Only that data are useful to Statistics which are homogeneous. An item of the data, like the price of a commodity, income of a farmer, etc. is called observation, or value, or measure, or item, or magnitude, etc. To draw any conclusion from



these data, an investigator has to first organise them. To do so, an investigator has to classify the same in the form of series.

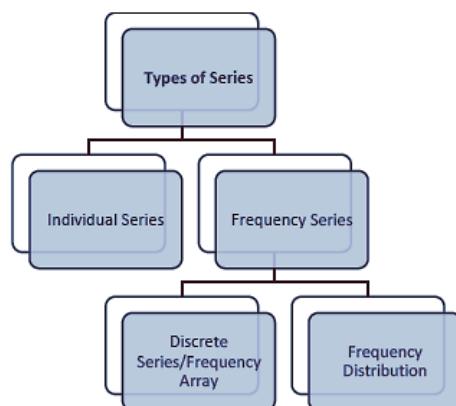
Series: Raw data are classified in the form of series. Series refer to those data which are presented in some order and sequence. Arranging of data in different classes according to a given order is called series. Thus, if the marks obtained by the students of Class XI are arranged according to their roll numbers in the ascending or descending order, the data so arranged would be known as statistical series. According to Horace Secrist, "4 series as used statistically may be defined as things or attributes of things arranged according to some logical order."

Univariate, Bivariate and Multivariate

- 'Uni' means one, 'bi' means two, 'multi' means many. Accordingly, univariate refers to a series of statistical data with one variable only, like the data on income of the households of a particular region.
- Bivariate refers to a series of statistical data with two variables like the data on income as well as expenditure of the households of a particular region.
- Multivariate refers to a series of statistical data with many (and more than two) variables, like the data on age, sex, education, income and expenditure of the households of a particular region.

6. Broadly, statistical series are of two types:

- (1) Individual Series or Series without Frequencies, and
- (2) Frequency Series or Series with Frequencies
Frequency series are further divided as:
 - (i) Discrete Series or Frequency Array, and
 - (ii) Frequency Distribution or Series with Class-Intervals.



(1) Individual Series: Individual series are those series in which the items are listed singly. For example, if the marks obtained by 30 students of Class XI are listed singly, the series would be called Individual Series. In these series there is no class of the items and also there is no frequency of the items. These series may be presented in two ways:

- (i) **According to Serial Numbers:** One way of presenting an individual series is that all the items are arranged in a serial order. Thus, marks obtained by the students may be arranged in order of their roll numbers. Data on the monthly expenses of the hostel students may be arranged in order of their room numbers. Data given in Table 4 on the marks obtained by 30 students are presented in Table 5 in order of their roll numbers.

Table : Marks Obtained by the Students in Statistics

Roll Number	Marks	Roll Number	Marks	Roll Number	Marks	Roll Number	Marks
1	30	9	20	17	15	25	35
2	25	10	10	18	30	26	20
3	15	11	40	19	20	27	20
4	30	12	20	20	15	28	25
5	25	13	10	21	15	29	35
6	20	14	30	22	25	30	10
7	10	15	25	23	20		
8	45	16	20	24	15		

- (ii) **Ascending or Descending Order of Data:** The other way of presenting an individual series is a simple ascending or descending order. In the ascending order, the smallest value is placed first, while in the descending order the highest value is placed first. Tables 6 and 7 show the arrangement of data in the ascending and descending orders respectively.

Table: Data Arranged in Ascending Order

10	15	20	25	30
10	15	20	25	30
10	15	20	25	55
10	20	20	25	35
15	20	20	30	40
15	20	25	30	45

Table : Data Arranged in Descending Order

45	30	25	20	15
40	30	20	20	15
35	25	20	20	10
35	25	20	15	10
30	25	20	15	10
30	25	20	15	10

Organisation of data in the form of individual series is a very simple form of presentation of data. But this method is not of much use when the number of items is very large.

(2) Frequency Series

Frequency series or series with frequencies may be of two types:

- (i) Discrete Series or Frequency Array, and
- (ii) Frequency Distribution.

Before we discuss these two types of series, let us understand the meaning of the following terms:

- (a) Frequency:** Frequency is the number of times an item occurs (or repeats itself) in the series. In other words, the number of times an item repeats itself in the population, is called the frequency of that item. For example, in Table 4, 10 has occurred 4 times. This means 4 students have secured 10 marks; or the frequency of 10 is 4.
- (b) Class Frequency:** The number of times an item repeats itself corresponding to a range of value (or class interval) is called class frequency. For example, if there are 4 students securing marks between 10-15, then 4 is the frequency corresponding to the class interval 10-15. Thus, 4 will be called class frequency.
- (c) Tally Bars:** Every time an item occurs, a tally bar, (|) is marked against that item. Corresponding to a particular class interval, each tally bar signifies 'one' occurrence of that item. Two tally bars would mean that the concerned item has occurred twice in the series. After every four tallies the fifth tally will cross out all the previous four tallies. Thus, making a group of five, i.e., [“H]- This method of marking and counting is known as Four and Cross Method. To

illustrate, Table 5 shows that 4 students obtained 10 marks, 5 students obtained 15 marks, 8 students obtained 20 marks, 5 students obtained 25 marks, 4 students obtained 30 marks, 2 students obtained 35 marks, 1 student obtained 40 marks and 1 student obtained 45 marks. All these frequencies have been presented in Table 8, using Four and Cross Method.

Table : Four and Cross Method of converting Raw Data into Frequency Series (data as in Table 5 or 6)

Marks	Tally Bars	Frequency
10		4
15		5
20		8
25		5
30		4
35		2
40		1
45		1

In this table, to express 8 tally bars, first of all four tally bars (||||) are marked, fifth tally bar has been marked across the four ||||. The sign |||| signifies that an item occurs five times in the series. Likewise, three tally bars are further marked |||| to make it equal to eight, i.e., |||| + |||| = 8. Thus, 1 = |; 2 = ||; 3 = |||; 4 = ||||; 5 = |||||; 6 = |||||; 7 = |||||; and so on.

- (i) Discrete Series or Frequency Array:** A discrete series or frequency array is that series in which data are presented, in a way that exact measurements of items are clearly shown. In such series there are no class intervals, and a particular item in the series is numbered rather than measured with some range.
- (ii) Frequency Distribution:** It is that series in which items cannot be exactly measured. The items assume a range of values and are placed within the range or limits. In other words, data are classified into different classes with a range, the range is called class intervals. Each item in the series is written against a particular class interval by way of a tally bar. The number of times an item occurs is shown as frequency against the class intervals to which that item belongs.



Marks	Tally Bars	Frequency
10-15		4
15-20		5
20-25		8
25-30		5
30-35		4
35-40		2
40-45		1
45-50		1

It is clear from above table that frequency of class interval 10-15 is 4. It means that there are 4 students who have secured marks between 10-15. Likewise, frequency of class interval 20-25 is 8 which means that there are 8 students who have secured marks between 20-25. But, it is not clear that how many students have secured 10 marks in the class interval 10-15 and how* many have secured 11 and 14 marks in the same class interval.

Size of Class: 'Size of Class' refers to size of the class interval, or it refers to width of the class. If 'range' (the difference between highest value and lowest value of the series) is (say) 100 and the number of classes is 20, then size of the class will be $100/20 = 5$. Thus: Size of the Class (S) = Range (r)/No. of Classes (n)

or $S=r/n = 100/20 = 5$ considering the above example.

Note; Size of the class must be such that all values belonging to the particular class interval tend to converge on the mid-value of the class interval. Only then it becomes an ideal class-size. Otherwise, our result would have a high degree of statistical error.

Some Important Terms

Let us understand some important terms before a detailed study of different types of frequency distribution in the next section.

- (i) **Class:** A range of values which incorporate a set of items is called a class. For example, 5-10, 10-15 are the classes.
- (ii) **Class Limits:** The extreme values of a class are limits. Every class interval has two limits, lower limit and upper limit. Of the class interval 5-10 in the above example, the lower limit is 5 and the upper limit is 70.

- (iii) **Magnitude of a Class Interval:** Magnitude of a class interval is the difference between the upper limit and the lower limit of a class. For example, in a class interval 10-15, the magnitude of the class interval would be $15 - 10 = 5$. Thus, Magnitude of a Class Interval (i) = Upper limit (l₂) - Lower limit (l₁)
The following formula is used to find out Class Interval.

Formula

$$i = l_2 - l_1$$

Where, i - magnitude of a class interval

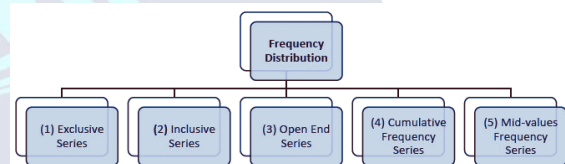
l₂ = upper limit of the class interval

l₁ = lower limit of the class interval.

- (iv) **Mid-value:** Mid-value is the average value of the upper and lower limits. It is known by adding up the upper limit and lower limit values and dividing the total by 2. Thus, Mid-value = $\frac{\text{Upper Limit} + \text{Lower Limit}}{2}$ where, m = mid-value; l₁ = lower limit; l₂ = upper limit.

For example, mid-value of 10-20 class interval $20+10/2 = 15$

7. Frequency Distribution is of five types:



- (1) **Exclusive Series:** Exclusive series is that series in which every class interval excludes items corresponding to its upper limit. In this series the upper limit of one class interval is the lower limit of the next class interval. It is called exclusive series because frequencies of the upper limit of each class interval is not included in that class.

For example, in a class interval, 10-15, only such items would be included, the value of which is between 10 and 14. Any item of the value of 15 would be included in the next class interval, viz., 15-20.

- (2) **Inclusive Series:** An inclusive series is that series which includes all items upto its upper limit. In such series, the upper limit of class interval does not repeat itself as a lower limit of the next class interval. Thus, there is a gap between the upper limit of a class interval and the lower limit of the next class interval. The gap ranges between 0.1 to 1.0. For



4

Presentation of Data

Important terms and concepts

- ❖ **Tabulation:** Orderly arrangement of data in rows and columns.
- ❖ **Objectives of Tabulation:**
 - Helps in understanding and interpreting the data easily.
 - It helps in comparing data.
 - It saves space and time.
 - Tabulated data can be easily presented in the form of diagrams and graphs.
- ❖ **Main parts of a table:**
 - **Title of the table:** It is a brief explanation of contents of the table.
 - **Table number:** It is given to be used for reference.
 - **Captions:** A word or phrase which explains the content of a column of a table.
 - **Stubs:** Stubs explain contents of row of a table.
 - **Body of the table:** Most important part of table as it contains data.
 - **Head note:** Head note is inserted to convey complete information of title.
 - **Source note:** Refers to the source from which information has been taken.
 - **Foot note:** It is used for pointing exceptions to the data

Types of Table -

- ❖ **Simple Table:** Data are presented according to one characteristic only.
- ❖ **Double Table:** Data are presented about two interrelated characteristics of a particular variable.
- ❖ **Three way table:** This table gives information regarding three interrelated characteristics of a particular variable.
- ❖ **Manifold table:** This table explains more than three characteristics of the data.

Diagrammatic Presentation of Data -

- ❖ **Utility or uses of diagrammatic presentation:**
 - Makes complex data simple.
 - Diagrams are attractive.
 - Diagrams save time when compared to other methods.
 - Diagrams create a lasting impression on the minds of observers.

Limitations of diagrammatic presentation -

- ❖ They do not provide detailed information.
- ❖ Diagrams can be easily misinterpreted.
- ❖ Diagrams can take much time and labour.
- ❖ Exact measurement is not possible in diagrams.

Kinds of diagrams -

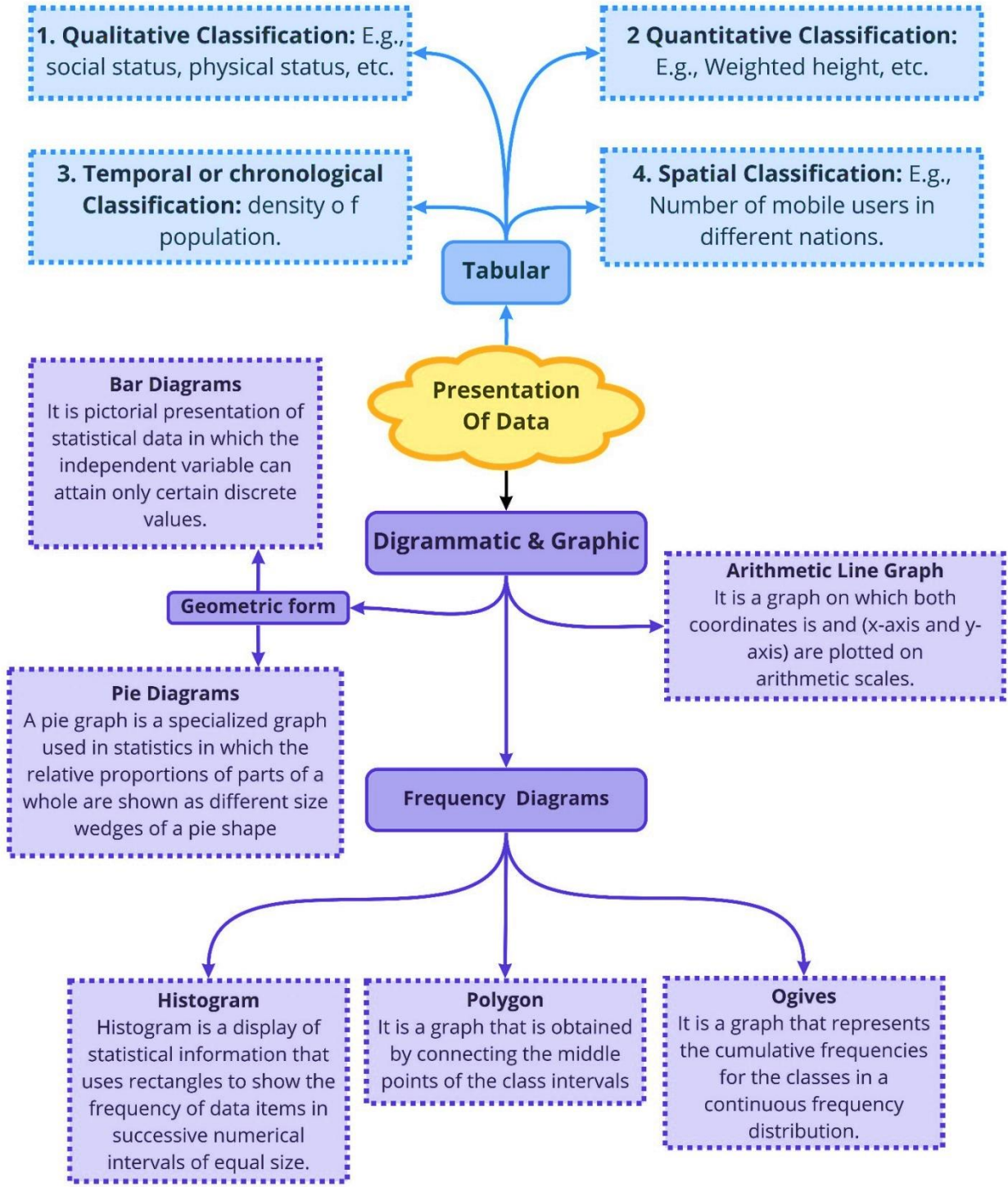
- ❖ **Line diagrams:** Lines are drawn vertically to show large number of items.
- ❖ **Bar diagram:**
 - **Simple Bar diagrams:** These diagrams represent only one particular type of data.
 - **Multiple Bar diagrams:** These diagrams represent more than one type of data at a time.
 - **Subdivided Bar diagram or Component Bar diagram:** These diagrams present total values and parts in a set of a data.
- ❖ **Pie diagrams:** Circle may be divided into various sectors representing various components.

GRAPHIC PRESENTATION OF DATA -

- ❖ **Advantages of Graphic Presentation:**
 - Graphs represent complex data in a simple form.
 - Values of median, mode can be found through graphs.
 - Graphs create long lasting effect on people's mind.
- ❖ **Disadvantages of graphic Presentation:**
 - Graphs do not show precise values.
 - Only experts can interpret graphs.
 - Graphs may suggest wrong conclusions.
- ❖ **Rules of Constructing graph:**
 - The heading of the graph should be simple, clear and self explanatory.
 - Graphs should always be drawn with reference to some scale.
 - False baselines should be drawn if the difference between zero and the smallest value is high.
 - Index should be made if different lines are drawn as in time series graphs.
- ❖ **Types of Graphs:**
 - **Line frequency graphs:** Such graphs are used to represent discrete series.
 - **Histogram:** A two dimensional diagram whose length shows frequency and the breadth shows size of class interval.
 - **Frequency Polygon:** A histogram becomes frequency polygon when a line is drawn joining midpoints of tops of all rectangles in a histogram. **Frequency Curve:** Smooth curve joining the points corresponding to the frequency and provides frequency curve of the data.
 - **Ogive:** A curve obtained by plotting frequency data on the graph paper.



Class : 11th Economics (Statistics)
Chapter-4 : Presentation Of Data



Important Questions

Multiple Choice Questions-

- The process of presenting data in the form of a table is called:
 - organisation
 - classification
 - presentation
 - tabulation
- Series in which frequencies are continuously added corresponding to each class interval in the series
 - Cumulative frequency series
 - Frequency
 - Deviation
 - Mid value
- 'Stub' of a table is the
 - Right part of the table describing the columns
 - Left part of the table describing the columns
 - Right part of the table describing the rows
 - Left part of the table describing the rows
- Classification data based on the geographical differences of the data is
 - Qualitative
 - Quantitative
 - Spatial
 - Chronological
- Whether classification is done first or tabulation`
 - Classification follows tabulation
 - Classification precedes tabulation
 - Both are done simultaneously
 - No criterion
- Arrangement of data according to some logical order is termed as
 - Collection series
 - Sample series
 - Statistical series
 - None
- The principal component of a table is:
 - table number
 - title
 - head note
 - all of these
- Which of the following is a basis of classification of a table?
 - Purpose
 - Construction
 - Originality
 - All of these
- The frequency distribution of two variables is known as
 - Univariate distribution
 - Sub- multivariate distribution
 - Bivariate distribution
 - Multivariate distribution
- Complex table may be classified as:
 - general purpose and special purpose table
 - original and derived
 - double, treble and manifold table
 - none of these
- In temporal classification, data are classified on the basis of:
 - location
 - time
 - originality
 - purpose
- Sub-divided bar diagram is used to:
 - Study relation between different components
 - Compare different components of a variable
 - Either (a) or (b)
 - Both (a) and (b)
- In tabulation, source of the data, if any, is shown in the:
 - Source Note
 - Body
 - Stub
 - Caption
- Series of statistical data with one variable only is called
 - None
 - Discrete
 - Continuous
 - Individual Series



15. The most accurate mode of data presentation is:
- Diagrammatic method
 - Tabulation
 - Textual presentation
 - None of these
2. In these questions, a statement of assertion followed by a statement of reason is given. Choose the correct answer out of the following choices.
- Both Assertion (A) and Reason (R) are true, and Reason (R) is the correct explanation of Assertion (A)
 - Both Assertion (A) and Reason (R) are true, and Reason (R) is not the correct explanation of Assertion (A)
 - Assertion (A) is true, but Reason (R) is false.
 - Assertion(A) is false, but Reason (R) is true.

Very Short Questions:

- Define presentation of data.
- What is tabulation?
- Define table.
- Explain a simple table.
- What is a complex table?
- Define a derived table.
- What are the two principal parts of a table?
- What are the different kinds of a table?
- What are the two types of complex table?
- What are a different form of presentation of data?

Short & Long Questions:

- Textual Presentation?
- Tabular Presentation?
- Components of a Table?
- Guidelines for the construction of a table or features of a good table?
- Kinds of Tables?
- Classification of Data and Tabular Presentation?
- Merits of Tabular Presentation?

Assertion Reason Question:

- In these questions, a statement of assertion followed by a statement of reason is given. Choose the correct answer out of the following choices.
 - Both Assertion (A) and Reason (R) are true, and Reason (R) is the correct explanation of Assertion (A)
 - Both Assertion (A) and Reason (R) are true, and Reason (R) is not the correct explanation of Assertion (A)
 - Assertion (A) is true, but Reason (R) is false.
 - Assertion(A) is false, but Reason (R) is true.

Assertion (A): In histogram no space is left between consecutive rectangles.

Reason (R): Histogram is a Graphical presentation of data and it is drawn for only continuous series.

Case Study Question:

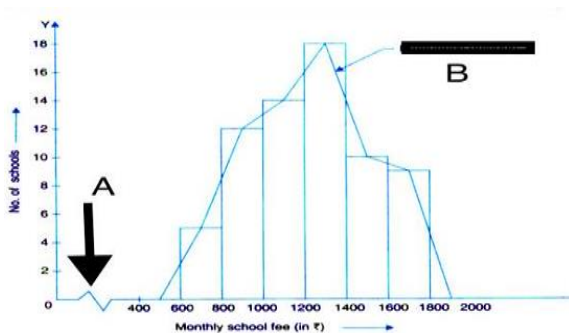
- 50 students were asked to choose their Favorite sport these are the results.

Sport	Cricket	Football	Hockey	Basketball	Tennis
No. of Student	12	10	11	9	8

- What angle should be used for football.
 - 36°
 - 72°
 - 90°
 - 10°
- In which form a data presented in a pie diagram?
 - percentage
 - Degrees
 - Absolute values
 - table
- In a pie diagram/circle 1% is equal to.....
 - 1
 - 3.6
 - 36
 - 10
- What angle should be used for tennis.
 - 8°
 - 57.6°
 - 28.8°
 - 64°

2. Answer the following questions based on the histogram Given below.

Monthly School Fee ₹	No. of Schools
600-800	5
800-1000	12
1000-1200	14
1200-1400	18
1400-1600	10
1600-1800	9



1. What A denotes in the above diagram.
 - (a) Jagged line
 - (b) horizontal line

- (c) Broken line
- (d) Both A and C

2. Which group contains maximum monthly school fee?
 - (a) 1000-1200
 - (b) 1200-1400
 - (c) 1400-1600
 - (d) 800-1000
3. How many schools charge monthly school fee between 1200 to 1800?
 - (a) 44
 - (b) 37
 - (c) 42
 - (d) 18
4. A _____ becomes a _____ if we Draw a line joining mid points of the tops of all rectangular.
 - (a) Histogram, Frequency polygon
 - (b) Frequency polygon, Histogram
 - (c) Frequency, Histogram
 - (d) Histogram, Frequency distribution

ANSWER KEY

Multiple Choice Answers-

1. D
2. A
3. D
4. C
5. B
6. C
7. D
8. D
9. C
10. C
11. B
12. D
13. A
14. D
15. B

Very Short Answers:

1. The presentation of data is a representation of data in an attractively and transparent manner that everybody understands and analyses.

2. Tabulation means presenting data in tabular form.
3. Table refers to the systematic representation of data with rows and columns.
4. A simple table only displays one characteristic of the data.
5. It is a table that shows more than one characteristic of the data.
6. A derived data is something where the information is not displayed according to the way it was collected. It is first into ratios or percentages and then presented.
7. The two principal parts of a table are (i) Table number and (ii) Title
8. Tables can be classified into three parts. (i) Purpose (ii) Originality, and (iii) Construction
9. The two types of a complex table are (i) Double or two-way table and (ii) Manifold table
10. The different form of presentation of data are (i) Textual and descriptive presentation (ii) Tabular presentation, and (iii) Diagrammatic presentation



Short & Long Answers:

1. In textual presentation, data are a part of the text of study or a part of the description of the subject matter of study. Such a presentation is also called descriptive presentation of data. This is the most common form of data presentation when the quantity of data is not very large. Here are some examples:

Example 1

In a strike call given by the trade unions of shoe making industry in the city of Delhi, 50% of the workers reported for the duty, and only 2 out of the 20 industries in the city were totally closed.

Example 2

Surveys conducted by a Non-government Organisation reveal that, in the state of Punjab, area under pulses has tended to shrink by 40% while the area under rice and wheat has tended to expand by 20%, between the years 2001-2011.

Suitability

Textual presentation of data is most suitable when the quantum of data is not very large. A small volume of data presented as a part of the subject matter of study becomes a useful supportive evidence to the text. Thus, rather than saying that price of gold is skyrocketing, a statement like price of gold has risen by 50% during the financial year 2017-18 is much more meaningful and precise. One need not support the text with voluminous data in the form of tables or diagram when the textual matter itself is very small and includes only a few observations. Indeed, textual presentation of data is an integral component of a small quantitative description of a phenomenon. It gives an emphasis of statistical truth to the otherwise qualitative observations.

Drawbacks

A serious drawback of the textual presentation of data is that one has to go through the entire text before quantitative facts about a phenomenon become evident. A picture or a set of bars showing increase in the price of gold during a specified period is certainly quite informative even on a casual glance of the reader. Textual presentation of data, on the other hand, does not offer anything to the reader at a mere glance of the text matter. The reader must read and comprehend the entire text. When the subject under study is vast and involves comparison across different areas/countries, textual presentation of data would only add to discomfort of the reader.

2. In the words of Neiswanger, "A statistical table is a systematic organisation of data in columns and rows" Vertical dissections of table (|) are known as columns and horizontal dissections (=) are known as rows.

Tabulation is the process of presenting data in the form of a table. According to Prof. L.R. Connor, 'tabulation involves the orderly and systematic presentation of numerical data in a form designed to elucidate the problem under consideration.'

In the words of Prof. M.M. Blair, "Tabulation in its broadest sense is an orderly arrangement of data in columns and rows."

3. Following are the principal components of a table:

(1) Table Number: First of all, a table must be numbered. Different tables must have different numbers, e.g., 1, 2, 3, etc. These numbers must be in the same order as the tables. Numbers facilitate location of the tables.

(2) Title: A table must have a title. Title must be written in bold letters. It should attract the attention of the readers. The title must be simple, clear and short. A good title must reveal:

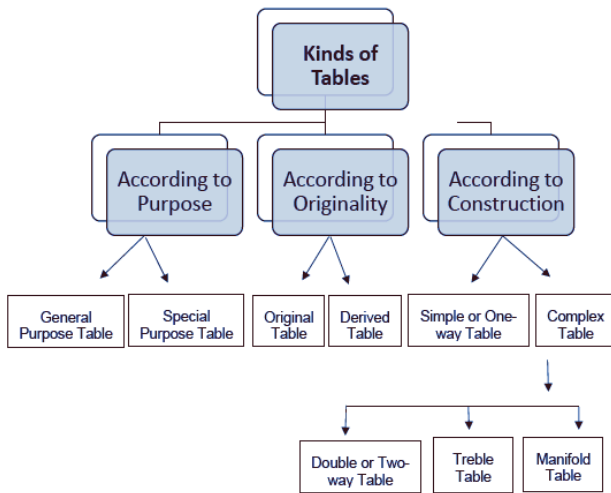
- (i) the problem under consideration,
- (ii) the time period of the study,
- (iii) the place of study, and

(iv) the nature of classification of data. A good title is short but complete in all respects.

(3) Head Note: If the title of the table does not give complete information, it is supplemented with a head note. Head note completes the information in the title of the table. Thus, units of the data are generally expressed in the form of lakhs, tonnes, etc. and preferably in brackets as a head-note.

(4) Stubs: Stubs are titles of the rows of a table. These titles indicate information contained in the rows of the table.

(5) Caption: Caption is the title given to the columns of a table. A caption indicates information contained in the columns of the table. A caption may have sub-heads when information contained in the columns is divided in more than one class. For example, a caption of 'Students' may have boys and girls as sub-heads.



Let us attempt a brief description of the various kinds of tables:

(1) Tables according to Purpose: According to purpose, there are two kinds of tables:

(i) General Purpose Table: General purpose table is that table which is of general use. It does not serve any specific purpose or specific problem under consideration. Such tables are just 'data bank' for the use of researchers for their various studies. These tables are generally attached to some official reports, like Census Reports of India. These are also called Reference Tables.

(ii) Special Purpose Table: Special purpose table is that table which is prepared with some specific purpose in mind. Generally, these are small tables limited to the problem under consideration. In these tables data are presented in the form of result of the analysis. That is why these tables are also called summary tables.

(2) Tables according to Originality: On the basis of originality, tables are of two kinds: (i) Original Table: An original table is that in which data are presented in the same form and manner in which they are collected. (ii) Derived Table: A derived table is that in which data are not presented in the form or manner in which these are collected. Instead the data are first converted into ratios or percentage and then presented.

(3) Tables according to Construction: According to construction, tables are of two kinds:

(i) Simple or One-way Table: A simple table is that which shows only one characteristic of the data. Table 2 below is an example of a simple table. It shows number of students in a college:

Number of Students in a College

Class	Number of Students
XI	200
B.A. (I)	100
B.A. (III)	80
B.A. (III)	60
Total	440

(ii) Complex Table: A complex table is one which shows more than one characteristic of the data. On the basis of the characteristics shown, these tables may be further classified as:

(a) Double or Two-way Table: A two-way table is that which shows two characteristics of the data. For example, Table 3, showing the number of students in different classes according to their sex, is a two-way table:

Number of Students in a College

(According to Sex and Class)

Class	Number of Students		Total
	Boys	Girls	
XI	160	40	200
B.A. (I)	40	60	100
B.A. (II)	60	20	80
B.A. (III)	50	10	60
Total	310	130	440

(b) Treble Table: A treble table is that which shows three characteristics of the data. For example, Table 4 shows number of students in a college according to class, sex and habitation.

Number of Students in a College

(According to Class, Sex and Habitation)

Class	Boys			Girls			Total		
	Rural	Urban	Total	Rural	Urban	Total	Rural	Urban	Total
XI	50	110	160	10	30	40	60	140	200
B.A. (I)	10	30	40	15	45	60	25	75	100
B.A. (IT)	15	45	60	5	15	20	20	60	80
B.A. (III)	10	40	50	5	5	10	15	45	60
Total	85	225	310	35	95	130	120	320	440

(c) Manifold Table: A manifold table is the one which shows more than three characteristics of the data. Table 5, for example, shows number of students in

a college according to their sex, class, habitation and marital status.

Number of Students in a College

(According to their Sex, Class, Habitation and Marital Status)

Class	Boys				Girls				Total
	Rural		Urban		Rural		Urban		
	Married	Unmarried	Married	Unmarried	Married	Unmarried	Married	Unmarried	
XI	5	55	10	90	2	8	5	25	200
B.A. (I)	5	15	15	35	4	4	4	18	100
B.A. (II)	5	10	15	30	2	3	5	10	80
B.A. (III)	5	5	20	20	3	2	2	3	60
Total	20	85	60	175	11	17	16	56	440

6. Tabular presentation is based on four-fold classification of data, viz., qualitative, quantitative, temporal, and spatial. Following are the details with suitable illustrations.

(1) Qualitative Classification of Data and Tabular Presentation:

Qualitative classification occurs when data are classified on the basis of qualitative attributes or qualitative characteristics of a phenomenon. Example: Data of unemployment may relate to rural-urban areas, skilled and unskilled workers, or male and female job-seekers. Table 6 below is an example of tabular presentation of data when data are classified on the basis of qualitative attributes or qualitative characteristics.

Unemployment in Punjab by Sex and Location (Per cent)

Sex	Location	
	Rural	Urban
Male	20	10
Female	30	20
Total	25	15

(This is an imaginary table. In this table, male and female are such characteristics/attributes which are qualitative and cannot be quantified.)

(2) Quantitative Classification of Data and Tabular Presentation:

Quantitative classification occurs when data are classified

on the basis of quantitative characteristics of a phenomenon.

Example: Data on marks in Mathematics by the students of Class XII in CBSE examination. Table 7 shows tabular presentation of data when data are classified on the basis of quantitative characteristics.

Marks Obtained by Students of Class XII of XYZ School

Marks	Number of Students
20-30	3
30-40	7
40-50	12
50-60	22
60-70	32
70-80	36
80-90	09
90-100	01

Source: Result Sheets Here, marks are a quantifiable variable and data are classified in terms of different class intervals of marks.

(3) Temporal Classification of Data and Tabular Presentation:

In temporal classification, data are classified according to time, and time becomes the classifying variable.

Example: Sale of Cell phones in different years during the period 2014-2018 in the city of Delhi. Table 8 shows tabular presentation of data on the basis of temporal classification.

Annual Sale of Cell Phones in the City of Delhi (2014-2018)

Year	Sale (Units)
2014	50,000
2015	70,000
2016	90,000
2017	1,00,000
2018	2,00,000

(4) **Spatial Classification:** In spatial classification, place/location becomes the classifying variable. It may be a village, a town, a district, a state or a country as a whole.

Example: Number of Indian students studying in different countries of the world during a particular year. Table 9 is an example of tabular presentation based on spatial classification of data.



Indian Students in different Countries of the World (2018)

Country	Number of Student
USA	50,000
UK	15,000
JAPAN	5,000
Russia	2,000
Australia	7,000

7. Following are the principal merits of tabular presentation of data:

- (1) **Simple and Brief Presentation:** Tabular presentation is perhaps the most simplest form of data presentation. Data, therefore, are easily understood. Also, a large volume of statistical data is presented in a very brief form.
- (2) **Facilitates Comparison:** The tabulation facilitates comparison of data by presenting the data in different classes.
- (3) **Easy Analysis:** It is very easy to analyse the data from tables. It is by organising the data

in the form of table that one finds out their central tendency, dispersion and correlation.

(4) **Highlights Characteristics of Data:** Tabulation highlights characteristics of data.

Accordingly, it becomes easy to remember the statistical facts.

(5) **Economical:** Tabular presentation is a very economical mode of data presentation. It saves time as well as space.

Assertion Reason Answer:

1. (a) Both Assertion (A) and Reason (R) are true, and Reason (R) is the correct explanation of Assertion (A).
2. (b) Both Assertion (A) and Reason (R) are true and Reason (R) is not the correct explanation of Assertion (A).

Case Study Answer:

1. **Answer:**
Q.1-b, Q.2-a, Q.3-b, Q.4-b
2. **Answer:**
Q.1-d, Q.2-b, Q.3-b, Q.4-a



STEP UP
ACADEMY

5

Measures of Central Tendency

Important terms and concepts

Average: It is a value which is typical or representative of a set of data. Averages are also called Measures of Central Tendency.

❖ **Functions of Average:**

- Presents complex data in a simple form.
- Facilitates comparison.
- Helps government to form policies.
- Useful in Economic analysis.
- **Essentials of a good Average:**
- Simple to calculate.
- It should be easy to understand.
- Rigidly defined.
- Based on all items of observation.
- Least affected by extreme values.
- Capable of further algebraic treatment.
- Least affected by sampling fluctuation.
- Graphic measurement possible.

❖ **Types of Averages:**

- Arithmetic Mean
- Median
- Mode
- Quartiles

❖ **Arithmetic Mean (X):**

- It is the most common type of measures of central tendency.
- It is obtained by dividing the sum of all observation in a series by the total number of observation.

❖ **Calculation of Arithmetic Mean:**

Types of Series	Direct Method	Shortcut Methods	Step deviation Methods
Individual Series	$\bar{X} = \frac{\sum X}{N}$	$\bar{X} = A + \frac{\sum d}{N}$	$\bar{X} = A + \frac{\sum d'}{N} \times C$
Discrete series	$\bar{X} = \frac{\sum fx}{N}$	$\bar{X} = A + \frac{\sum fd'}{N}$	$\bar{X} = A + \frac{\sum fd'}{N} \times C$
Continuous Series	$\bar{X} = \frac{\sum fm}{N}$	$\bar{X} = A + \frac{\sum fd'}{N}$	$\bar{X} = A + \frac{\sum fd'}{N} \times C$



❖ **Merits of Arithmetic Mean:**

- Easy to calculate.
- Simple to understand.
- Based on all observations.
- Capable of further mathematical calculations.

Demerits:

- Affected by extreme values.
- Cannot be calculated in open-end series.
- Cannot be graphically ascertained.
- Sometimes misleading or absurd result.

❖ **Weighted Arithmetic Mean:**

Values to be arranged are given varying importance.

$$XW = \frac{\sum WX}{\sum W}$$

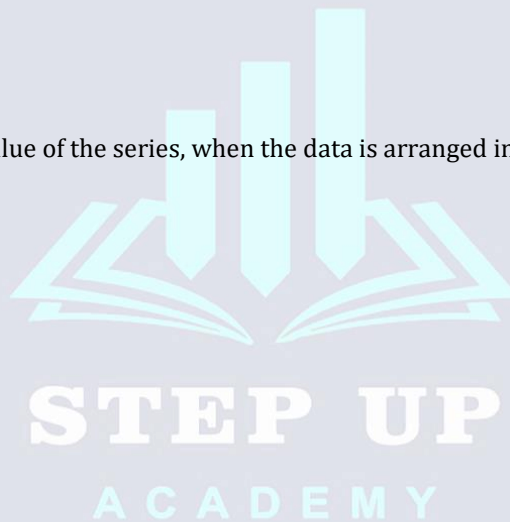
Where X_w = Weighted Arithmetic Mean

W = Weight

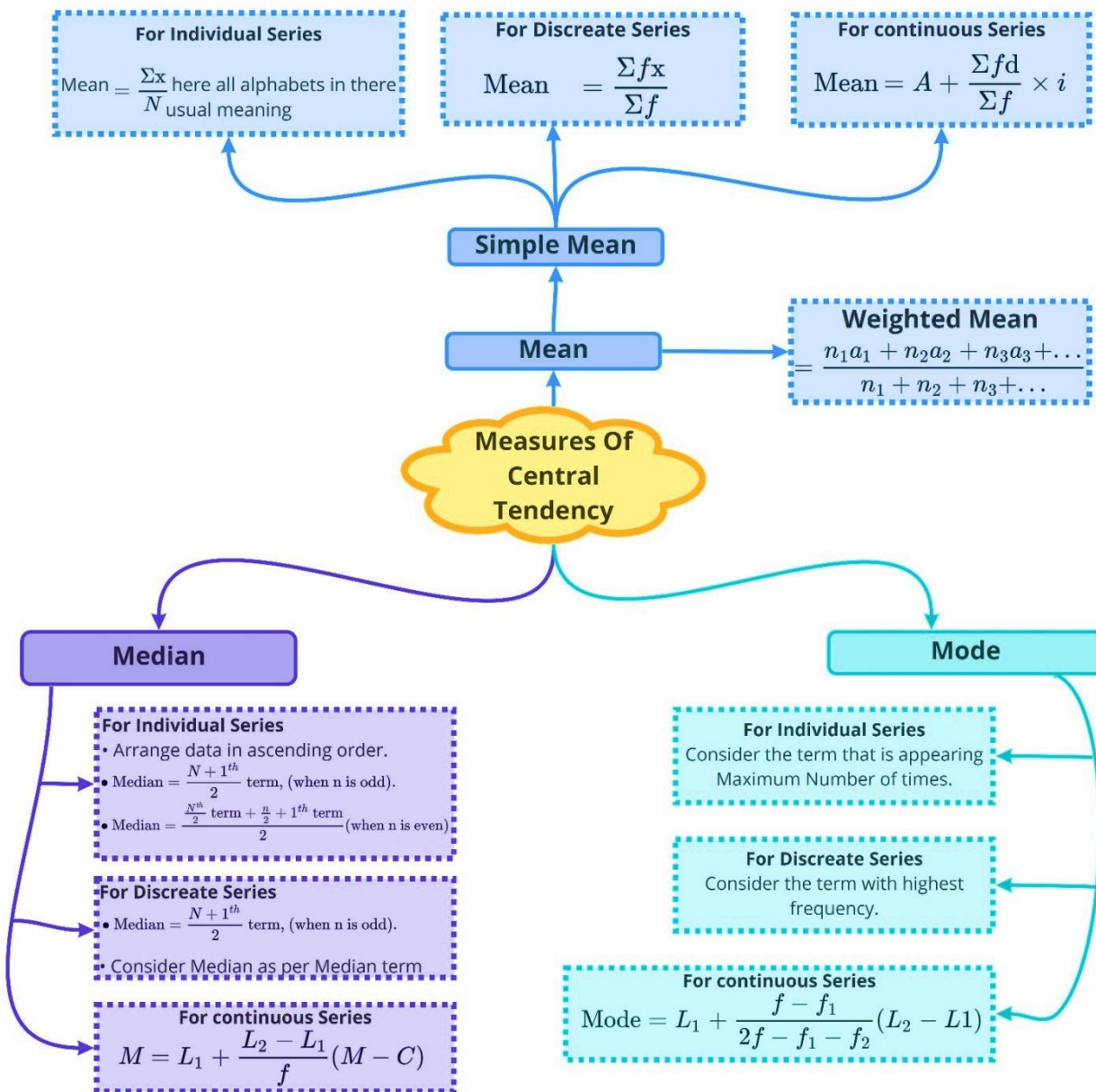
X = Values of the Variables

❖ **Median (M):**

It is defined as the middle value of the series, when the data is arranged in ascending or descending order.



Class : 11th Economics (Statistics)
Chapter-5 : Measures Of Central Tendency





Important Questions

Multiple Choice Questions-

- Which of the following is a type of mathematical average?
 - Median
 - Partition value
 - Mode
 - None of these
- Arithmetic mean of these items 5, 7, 9, 15, 20 is:
 - 10
 - 10.2
 - 11.2
 - 12
- Arithmetic mean of these items: 10, 15, X, 20, 30 is 20. Find out the missing item.
 - 10
 - 15
 - 5
 - 25
- While computing the arithmetic mean of a frequency distribution, the each value of a class is considered equal to:
 - Class mark
 - Lower limit
 - Upper limit
 - Lower class boundary
- Arithmetic mean of a series is 15 and if 5 is added in all the items of this series, the new arithmetic mean will be:
 - 5
 - 20
 - 18
 - (d)10
- Which of the following is not a measure of central tendency?
 - Mean
 - Mode
 - Standard deviation
 - Median
- Which is not a method to find arithmetic mean?
 - Direct method
 - Short-cut method
 - Step-deviation method
 - Karl Pearson's method
- The sample mean is a:
 - Parameter
 - Statistic
 - Variable
 - Constant
- Assumed mean is taken in which method?
 - Direct method
 - Step-deviation method
 - Karl Pearson's method
 - Spearman's method
- The sum of deviations taken from mean is:
 - Always equal to zero
 - Some times equal to zero
 - Never equal to zero
 - Less than zero
- Sum of deviations of different values from arithmetic mean is always equal to:
 - zero
 - one
 - less than one
 - more than one
- Which of the following statements is always true?
 - The mean has an effect on extreme scores
 - The median has an effect on extreme scores
 - Extreme scores have an effect on the mean
 - Extreme scores have an effect on the median
- Mean of 0.3, 5, 6, 7, 9, 12, 0.6 is:
 - 4.9
 - 5.7
 - 5.6
 - None of these
- Simple average is sometimes called:
 - Unweighted average
 - Weighted average
 - Relative average
 - None of these

Very Short Questions:

- Define median.
- What is the mode?
- Define the partition value.

4. Explain quartile.
5. What is positional average?
6. Define the central tendency.
7. What are the purpose of average is the statistical method?
8. What are the different kinds of statistical average?
9. What are the two methods that can calculate the simple arithmetic mean in case of individual series?
10. What are the methods calculating simple arithmetic mean?

Short & Long Questions:

1. Definition Of Central Tendency?
 2. Purpose And Functions Of Averages?
 3. Essentials Of A Good Average?
 4. Types of statistical averages?
5. Arithmetic mean?
 6. Types of arithmetic mean?
 7. Pocket allowance of 10 students is Rs. 15, 20, 30, 22, 25, 18, 40, 50, 55 and 65. Find out the average pocket allowance.
 8. Following is the pocket allowance of 10 students. Find out arithmetic mean using Shortcut Method.

Pocket Allowance (Rs.)	15	20	30	22	25	18	40	50	55	65

9. Weighted Arithmetic Mean?
10. Combined Arithmetic Mean?
11. Properties of Arithmetic Mean (AM)?
12. The Principal Merit of Arithmetic Mean?
13. Calculation of Arithmetic Mean for Different Series: A Glance?

ANSWER KEY

Multiple Choice Answers-

1. D
2. A
3. C
4. D
5. B
6. B
7. C
8. D
9. B
10. B
11. A
12. A
13. C
14. B
15. A

Very Short Answers:

1. Median is a value located centrally of a series in such a way that half of the value of the series is above it and the other half is below.
 2. The mode is a value that frequently occurs in the series. Which means the modal value has the highest frequency in the series.
3. The value that divides the series into more than two parts is known as a partition value.
 4. The end value of the statistical series when divided into four parts is known as quartile.
 5. Positional average are those averages whose value is worked out on the basis of their position in the statistical series.
 6. All the methods of statistical analysis by which the average of the statistical series are analysed is known as a central tendency.
 7. The purpose of the average is the statistical method are
 - Brief description
 - Comparison
 - Formulation of policies
 - Statistical analysis
 - One value of all
 8. The different kinds of statistical average are.
 - Mathematical average
 - Positional average
 9. The two methods that can calculate the simple arithmetic mean in the case of individual series are.
 - Direct method
 - Short-cut method



10. The methods of calculating simple arithmetic mean are.

- Individual series
- Discrete series
- Frequency distribution

Short & Long Answers:

1. A Central Tendency refers to an average or a central value of a statistical series. It is difficult for anyone to understand or remember a large group of raw data. One would like to know the critical value which represents all the items in a series. Such a value is called 'central tendency' or 'average value'. For instance it is very difficult to remember and understand the data concerning the income of millions of Indians. However, if it is said that in 2016-17, provisional estimates for average income of the people in India was Rs. 82,269 per annum, it will be easy for us to guess the economic condition of most of the Indians. It is this average value which is called central tendency of the series. It is also called measure of location. Thus, measures of central tendency refer to all those methods of statistical analysis by which averages of the statistical series are worked out.

Definition

According to Croxton and Cowden, "An average is a single value within the range of the data that is used to represent all of the values in the series. Since an average is somewhere within the range of data, it is sometimes called measure of central value."

According to Clark, "An average is a figure that represents the whole group."

2. Study of averages is of central significance in statistical methods. That is why Bowley defines statistics as, "A Science of Averages".

What is the Basic Purpose of Finding an Average Value of a Series?

It is to identify such a value that represents characteristics of all the items in the series.

According to Moroney, "The purpose of an average is brief and simple representation of a group of individual values so that the brain may quickly grasp the general basis of the units of the group." Some of the main purposes and functions of averages are as under:

(1) **Brief Description:** The main purpose of an average is to present a brief description of the principal features of the raw data. As a result, data are easily understood.

(2) **Comparison:** Averages help in making comparison of different sets of data. For example, a comparison of the per capita income of India and USA shows that per capita income of India is much less than the per capita income of USA. Accordingly, it is concluded that India is a poor country.

(3) **Formulation of Policies:** Averages help in formulation of policies. For example, in India the per capita income is Rs. 82,269 per annum which is much less than many countries in the world. Accordingly, it becomes clear for the government to focus on such economic policies as are likely to increase per capita income.

(4) **Statistical Analysis:** Averages constitute the basis of statistical analysis. For example, if one knows the average marks secured by the students of a class in their different subjects, one can easily analyse the subjects in which the students are weak.

(5) **One Value for All:** Averages represent the universe or the mass of statistical data. One value represents all values of the series. Accordingly, conclusion can be drawn in respect of the universe as a whole.

3. A good and satisfactory average should have the following features:

1) **Clear and Stable Definition:** A good and a satisfactory average should be clear and stable in definition.

2) **Representative:** An average value should be representative of the entire mass of data. It should be based on all the observations of the series.

3) **Simplicity:** Simplicity is another essential feature of a good average. It must be so simple that it is easily worked out.

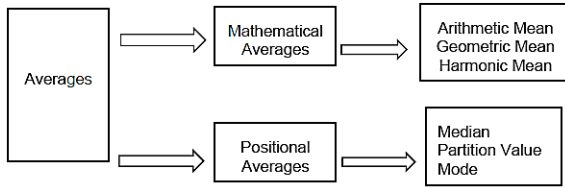
4) **Certainty:** A good average must be certain in character. Only then an average value can be used as the basis of statistical analysis.

5) **Absolute Number:** A good average should be an absolute number. A percentage or a relative value does not serve as a good average.

6) **Least Effect of a Change in the Sample:** An average of a series should be least affected by a change in the sample on which the average is based.

7) **Algebraic Treatment:** A good average should be capable of further mathematical or algebraic treatment.

4. TYPES OF STATISTICAL AVERAGES



Averages are broadly classified into two categories:

- (1) Mathematical Averages, and
- (2) Positional Averages.

In the present chapter we discuss arithmetic mean.

5. **ARITHMETIC MEAN-** Arithmetic Mean is a simple average of all items in a series. It is the simplest measure of central tendencies. The arithmetic mean of a series is simply called 'Mean'. If, for example, Ram plays 5 matches, Shyam 6, Mohan 7, Kishan 8 and Ravi 9 matches a week, the average number of matches played by Ram, Shyam, Mohan, Kishan and Ravi would be determined as under:

$$\text{Number of Matches} = 5 + 6 + 7 + 8 + 9 = 35$$

$$\text{Number of Boys} = 5$$

$$\text{Mean} = \frac{\text{Total Value of the Items}}{\text{Number of Items}} = \frac{35}{5} = 7$$

What is the basic difference between Simple Arithmetic Mean and Weighted Arithmetic Mean?

In simple arithmetic mean, all items of the series are taken as of equal importance. In the weighted average, on the other hand, different items are taken as of different importance; accordingly, weights are accorded to different items depending on their relative importance.

Definition

Arithmetic Mean or Mean is the number which is obtained by adding the values of all the items of a series and dividing the total by the number of items.

In the words of H. Secrist, "The Arithmetic Mean is the amount secured by dividing the sum of value of the items in a series by their numbers."

FORMULA

Arithmetic mean is generally written as X. It may be expressed in the form of following formula:

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{N} = \frac{\Sigma X}{N}$$

Here, X₁; X₂, X₃, X_n are the values of different items in the series. Thus, X₁ = matches played by Ram, 5; X₂ = matches played by Shyam, 6 and X₃ =

matches played by Mohan, 7, etc.

N = Total number of items (in the above example, it is 5 comprising of five boys, Ram, Shyam, Mohan, Kishan and Ravi).

Σ is a sign called Sigma. It refers to the sum total of the values of different items in the series.

6. Arithmetic mean is of two types:

(1) Simple Arithmetic Mean: In it, all items of a series are given equal importance.

(2) Weighted Arithmetic Mean: In it, different items of a series are accorded different weights in accordance with their relative importance.

Methods of Calculating Simple Arithmetic Mean

We know, there are three types of statistical series:

(1) Individual Series

(2) Discrete Series

(3) Frequency Distribution.

Arithmetic mean may be calculated with respect to these series using different methods. Calculation of Simple Arithmetic Mean in Case of Individual Series

In the case of individual series, arithmetic mean may be calculated by two methods:

(1) Direct Method

(2) Short-cut Method.

Direct Method

Following steps are involved in this method:

- (i) Add up values of all the items of a series (ΣX);
- (ii) Find out total number of items in the series (N); and

(iii) Divide the total of value of all the items (ΣX) with the number of items (N). The resultant value would be the arithmetic mean. Thus,

FORMULA

$$\bar{X} = \frac{\Sigma X}{N} \text{ OR } X = \frac{\text{TOTAL VALUE OF THE ITEMS}}{\text{NUMBER OF ITEMS}}$$

7. Average pocket allowance of the 10 students = Rs. 34.

Pocket Allowance (₹) (X)
15
20
30
22

series do not have frequencies of the items. These series show only values of different items.

In discrete series, there are frequencies corresponding to different items in the series.

There are three methods of calculating mean of the discrete series.

That discrete series (also called frequency array) do not have class intervals. An item in the series does not assume any range of values but each item has corresponding frequency.

- (1) Direct Method
- (2) Short-cut Method
- (3) Step-deviation Method.

(1) Direct Method

Direct method of calculating mean of the discrete series involves the following steps:

- (i) Values of the various items in the series are indicated by X, and their frequencies by 'f'.
- (ii) Each item is multiplied by its frequency to get 'fX'. These multiples are added to get ΣfX. That is, $\Sigma fX = f_1X_1 + f_2X_2 + \dots + f_nX_n$
- (iii) Frequencies are added up to get Σf. That is, $\Sigma f = f_1 + f_2 + f_3 + \dots + f_n$
- (iv) ΣfX is divided by Σf to obtain the mean, X.

FORMULA

$$\bar{X} = \Sigma fX / \Sigma f$$

9. We have so far discussed Simple Arithmetic Mean. In simple arithmetic mean, all items of a series are taken as of equal significance. But sometimes we may give greater significance to some items and less to others. As household may, for example, give more significance to food, less to clothes and still less to entertainment. When different items of a series are weighed according to their relative importance, the average of such series is called Weighted Arithmetic Average. A Weighted Arithmetic Average is, thus, the mean of weighted items.

Calculation of Weighted Mean

Calculation of weighted mean involves the following steps:

- (i) Different items are weighed according to their significance. Weights are indicated by 'W'.
- (ii) Items (X) are multiplied by their corresponding weights (W) and added up to get ΣWX.

- (iii) ΣWX is divided by the sum total of weights, i.e., ΣW, to get the mean value, that is,

FORMULA

$$\bar{X}_w = \Sigma WX / \Sigma W$$

Here, \bar{X}

w indicates weighted average.

10. Given the mean values of two or more parts of a series and the number of items in each part, one can get Combined Arithmetic Mean, or mean of the series as a whole. The following formula is used for the estimation of combined arithmetic mean:

FORMULA

$$\bar{X}_{1,2} = \bar{X}_1/N_1 + \bar{X}_2/N_2 / N_1 + N_2$$

Here, $\bar{X}_{1,2}$ = Combined arithmetic mean of parts 1 and 2 of a series;

\bar{X}_1 = Arithmetic mean of part 1 of the series;

\bar{X}_2 = Arithmetic mean of part 2 of the series;

N1 = Number of items in part 1 of the series;

N2 = Number of items in part 2 of the series.

Likewise, when there are more than 2 parts of a series, the following formula is used to work out Combined Arithmetic Mean. *

FORMULA

$$\bar{X}_{1,2,3,\dots,n} = \bar{X}_1N_1 + \bar{X}_2N_2 + \dots + \bar{X}_nN_n / N_1 + N_2 + \dots + N_n$$

11. (i) The sum of deviations of the items from AM is always zero.
- (ii) The sum of squared deviations of the items from AM is minimum.
- (iii) If each item of a series is increased, decreased, multiplied or divided by some constant then AM also increases, decreases, multiplies or is divided by the same constant.
- (iv) The product of the AM and the number of items on which mean is based is equal to the sum of all given items.
- (v) If each item of the original series is replaced by the actual mean, then the sum of these substitutions will be equal to the sum of the individual items.

12. That it is based on all items of the series and is capable of further algebraic treatment.

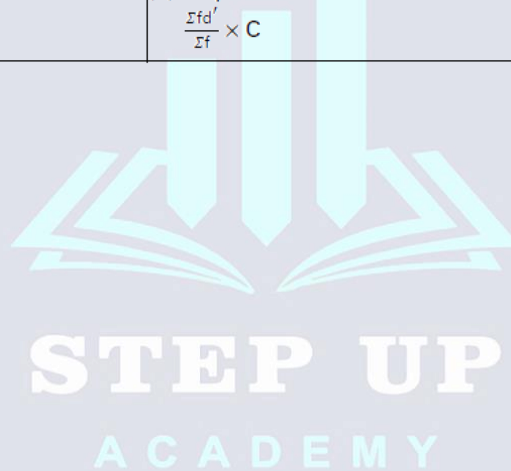
The Principal Demerit of Arithmetic Mean End-use of the goods is the principal basis of classifying the goods as intermediate goods and final goods.

Merits and Demerits of Arithmetic Mean



13.

Series	Methods of Calculation of Arithmetic Mean
1. Individual Series	(a) Direct Method: Formula: $\bar{X} = \frac{\sum X}{N}$
	(b) Short-cut Method: Formula: $\bar{X} = A + \frac{\sum d}{N}$
2. Discrete Series/Frequency Array	(a) Direct Method: Formula: $\bar{X} = \frac{\sum fX}{\sum f}$
	(b) Short-cut Method: Formula: $\bar{X} = A + \frac{\sum fd}{\sum f}$
	(c) Step-deviation Method: Formula: $\bar{X} = A + \frac{\sum fd'}{\sum f} \times C$
3. Frequency Distribution Series	(a) Direct Method: Formula: $\bar{X} = \frac{\sum fm}{\sum f}$
	(b) Short-cut Method: Formula: $\bar{X} = A + \frac{\sum fd}{\sum f}$
	(c) Step-deviation Method: Formula: $\bar{X} = A + \frac{\sum fd'}{\sum f} \times C$



6

Measures of Dispersion

Points to remember

Dispersion is a measure of the variation of the items from central value.

The measures of dispersion is important to compare uniformity, consistency and reliability amongst variables/ series.

Absolute measures of dispersion are expressed in terms of original unit of series.

Relative measures are expressed in ratios or percentage of average, also known as coefficients of dispersion.

Measures of Dispersion:

- Range
- Inter quartile range
- Quartile deviation or Semi-Inter-quartile range
- Mean deviation
- Standard Deviation
- Lorenz curve

Range: Range is defined as the difference between two extreme observations i.e. the largest and the smallest value.

Symbolically, $R = L - S$

Where, $R = \text{Range}$

$L = \text{Largest Value}$

$S = \text{Smallest value}$

Coefficient of range = $\frac{L-S}{L+S}$

Inter Quartile Range:

Inter quartile range is the difference between upper quartile and lower quartile.

Inter-quartile range = $Q_3 - Q_1$

Where, $Q_3 = \text{Third quartile or upper quartile.}$

$Q_1 = \text{First quartile or lower quartile}$

Quartile Deviation:

Quartile deviation is known as half of difference of third quartile (Q_3) and first quartile (Q_1). It is also known as semi inter quartile range.

$$Q.D. = \frac{Q_3 - Q_1}{2}$$

Where, Q.D. = Quartile deviation

$Q_3 = \text{Third quartile or upper quartile.}$

$Q_1 = \text{First quartile of lower quartile.}$

Coefficient of quartile deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$



Mean Deviation:

Mean deviation/average deviation is the arithmetic mean of the deviations of various items from their average (mean, median or mode) generally from the median.

Calculation of mean deviation

$$\text{Individual Series } M.D. = \frac{\sum |D|}{N}$$

$$\text{Discrete Series } M.D. = \frac{\sum f|D|}{N}$$

$$\text{Continuous Series } \frac{\sum |D|}{N}$$

Where, MD = Mean deviation

| D | = Deviations from mean or median ignoring + Signs

N = Number of item (Individual Series)

N = Total number of Frequencies (Discrete and continuous series)

F = Number of frequencies.

Coefficient of mean deviation

Merit of Mean deviation:

1. As in case of X, every term is taken in account hence, it is certainly a better measure than other measures of dispersion i.e. Range, Percentile Range or Quartile Range.
2. Mean deviation is extensively used in other fields such as Economics, Business, Commerce or any other field of such type.
3. It has least sampling fluctuations as compared to Range, Percentile Range and Quartile Deviation.
4. When comparison is needed this is perhaps the best measure between two or more series.
5. This calculation has its base upon measurement than an estimate.
6. Mean Deviation is rigidly defined; one of the main focus point of any measure used for statistical Analysis.
7. It we calculate it from median it is less affected by extreme terms.
8. As it is based on the deviations about an average, it gives us better measure for comparison.

Demerits of Mean Deviation:

1. If average is in fractions, it is difficult to compile M.D.
2. Main property is absent, It is not capable of further Algebraic Treatment.
3. Not so easy to calculate to calculate X, M or Z first and then to go for other measures.
4. If it is calculated from Z it is not much reliable as Mode (Z) is not the true representative of the series.
5. M.D. and its co-efficient taken from X, M and Z often differ.
6. As +, - signs are ignored which is not possible mathematically. Algebraically we have to proceed for Standard Deviation; or another measure of dispersion.
7. As for mean, open and series cannot be taken for the true result.
8. If Range increases in case the sample increases, Average deviation also increases but not in the same ratio.
9. For Sociological studies, it is almost not used.

Standard Deviation:

Standard deviation is the best and widely used measure of dispersion. Standard deviation is the square root of the arithmetic mean of the squares of deviation of its items from their arithmetic mean. Calculation of standard deviation in individual series.

Actual mean method.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

Where, σ = Standard Deviation

Σx^2 = Sum total of square of Deviation taken from Mean

N = Number of items

Shortcut Method or assumed mean method:

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$$

Where d^2 = Square of deviation taken from assumed mean.

Calculation of standard deviation in discrete series:

Actual mean method or direct method

$$\sigma = \sqrt{\frac{\sum x^2}{N}}$$

Where σ = S.D.

Σx^2 = Sum total of the squared deviations multiplied by frequency

N = Number of pair of observations.

Shortcut method or assumed method:

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$$

σ = S.D.

Σfd^2 = Sum total of the squared deviations Multiplied by frequency

Σfd = Sum total of deviations multiplied by frequency.

N = Number of pair of observations.

Step deviation method:

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2} \times C$$

σ = Standard Deviation

Σfd^2 = Sum total of the squared step deviations multiplied by frequency.

Σfd = Sum total of step deviations multiplied by frequency

C = Common factor

N = Number of pair of observation

Individual Series:

1. Actual Mean Method

$$\sigma = \sqrt{\frac{\sum x^2}{N}}$$

$$x = X - \bar{X}$$

2. Assumed Mean Method

$$\sigma = \sqrt{\frac{\sum d^2}{N} - \left(\frac{\sum d}{N}\right)^2}$$



Discrete/Continuous Series:

1. Actual Mean Method

$$\sigma = \sqrt{\frac{\sum fx^2}{\sum f}}$$

$$x = X - \bar{X}$$

2. Assumed Mean Method

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$$

3. Step Deviation Method

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2} \times C$$

Merits of standard deviation:

1. Based on all values
2. Rigidly defined
3. Less effect of fluctuations
4. Capable of algebraic treatment

Demerits of standard deviation:

1. Difficult to compute
2. More stress on extreme items
3. Dependent on unit of measurement

Coefficient of variation:

When two or more groups of similar data are to be compared with respect to stability (or uniformity or consistency or homogeneity). Coefficient of variation is the most appropriate measure. It is the ratio of the standard deviation to the mean.

$$CV = \frac{\sigma}{\bar{X}} \times 100$$

Where C.V. = Coefficient of variation

σ = Standard deviation

\bar{X} = Arithmetic mean

LORENZ CURVE:

The Lorenz curve devised by Dr. Max O. Lorenz is a graphic method of studying dispersion.

The Lorenz curve always lies below the line of equal distribution, unless the distribution is uniform.

The Area between the line of equal distribution and the plotted curve gives the extent of inequality in the items.

The larger the area, more is the inequality.

Application Lorenz Curve:

- (i) Distribution of income
- (ii) Distribution of wealth
- (iii) Distribution of wages
- (iv) Distribution of production
- (v) Distribution of population

Construction of Lorenz Curve:

1. Series is converted into a cumulative frequency series. The cumulative sum of items is assumed to be 100 and different items are converted into percentage of the cumulative sum.
2. Cumulative sum of frequency is assumed to be 100 and different Frequencies are converted into percentage of sum of frequency.
3. Cumulative frequencies are plotted on x-axis and cumulative items are plotted on y-axis of graph.
4. On both axis values are plotted from 0-100.
5. A diagonal line(0 on X-axis and 100 on Y-axis are joined by a line). It is called line of equal distribution.
6. Actual data are plotted by joining different points. This is Lorenz Curve.
7. Lesser distance between the line of equal distribution and line of actual distribution shows lesser dispersion and so on.





Class : 11th Economics (Statistics)
Chapter-6 : Measures Of Dispersion

Construction

- Items of both the groups are made cumulative and then cumulative frequencies are known,
- Considering the last cumulative frequency as 100 find the cumulative percentage of all frequencies,
- Now ascertain the points of the percentage of both the groups or both the axes of the curve,
- Draw a curve by joining both the ends with free hand and compare it with the line of equal distribution,
- In continuous series, we proceed by taking the midpoint of the group.

Meaning
The Lorenz curve is a kind of cumulative frequency curve which is based on percentage and it is drawn by putting values on graph.

- Coefficient of range = $\frac{L - S}{L + S}$
- Coefficient of quartile deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$
- Coefficient of mean deviation = $\frac{\text{mean deviation}}{\text{mean}}$
- Coefficient of standard deviation = $\frac{S.D.}{\text{mean}}$
- Coefficient of variation = $\frac{S.D.}{\text{mean}} \times 100$

Lorenz Curve

Relative Dispersion

Measures Of Dispersion

Rang
L - S

Absolute Dispersion

Quartile Deviation
Quartile Deviation = $\frac{Q_3 - Q_1}{2}$
Where $Q_1 = \frac{N^{th}}{4} \text{ term}$,
 $Q_3 = 3 \times \left(\frac{N^{th}}{4}\right) \text{ term}$

Mean Deviation

Standard Deviation

For continuous Series
 $M. D. = \frac{\sum f|d|}{N}$

For Discrete Series
 $M. D. = \frac{\sum f|d|}{N}$

For Individual Series
 $M. D. = \frac{\sum |d|}{N}$

For Individual Series
 $S. D. = \sqrt{\frac{\sum d^2}{N}}$

For assumed Mean Method
 $S. D. = \sqrt{\frac{\sum fd^2}{N} - \left[\frac{\sum fd}{N}\right]^2}$
Where d = deviation from assumed mean

For Discrete & continuous Series
 $S. D. = \sqrt{\frac{\sum fd^2}{N}}$

Important Questions

Multiple Choice Questions-

- When it comes to comparing two or more distributions we consider:
 - Absolute measures of dispersion
 - Relative measures of dispersion
 - Both (a) and (b)
 - Either (a) or (b)
- The most commonly used measure of dispersion is:
 - Range
 - Standard deviation
 - Coefficient of variation
 - Quartile deviation
- The range of 15, 12, 10, 9, 17, 20 is:
 - 5
 - 12
 - 13
 - 11
- The standard deviation of 10, 16, 10, 16, 10, 10, 16, 16 is:
 - 4
 - 6
 - 3
 - 0
- Corresponding to first quartile, the cumulative frequency is:
 - $N/2$
 - $N/4$
 - $3N/4$
 - None of these
- What is the coefficient of range for the following wages (in `) of 8 workers?
80, 65, 90, 60, 75, 70, 72, 85
 - 35
 - 25
 - 30
 - 20
- _____ divide the total number of observations into 4 equal parts.
 - median
 - deciles
 - quartiles
 - percentiles
- Above upper quartile, the frequency is equal to:
 - $N/4$
 - $N/2$
 - $3N/4$
 - None of these
- What is the value of mean deviation about mean for the following numbers?
5, 8, 6, 3, 4
 - 5.20
 - 7.20
 - 1.44
 - 2.23
- _____ is equal to the value corresponding to cumulative frequency $(N + 1)/4$ from simple frequency distribution.
 - Median
 - 1st quartile
 - 3rd quartile
 - 1st decile

Very Short Questions:

- Define dispersion.
- What is the coefficient of dispersion?
- Define range.
- Explain the interquartile range.
- What is the quartile deviation?
- What is the formula for calculating the coefficient of quartile deviation?

Short & Long Questions:

- Define mean deviation.
- What is standard deviation?
- What is a Lorenz curve?
- Define variance.
- A measure of dispersion is a good supplement to the central value in understanding a frequency distribution. Comment.
- Which measure of dispersion is the best and how?
- Some measures of dispersion depend upon the spread of values whereas some calculate the variation of values from a central value. Do you agree?
- In town, 25% of the persons earned more than ₹ 45,000 whereas 75% earned more than 18,000. Calculate the absolute and relative values of dispersion.



ANSWER KEY

Multiple Choice Answers-

1. B
2. B
3. D
4. C
5. B
6. D
7. C
8. A
9. C
10. B

Very Short Answers:

1. Dispersion is the measure of the extent to which different items tend to disperse away from the central tendency.
2. The coefficient of dispersion shows different data percentages or relative values. It is known as a relative measure of dispersion.
3. Range is the variance between the lowest and the highest values in a series.
Range = Highest value in the series - Lowest value in the series.
4. The difference between the first quartile (Q_1) and the third quartile (Q_3) of a series is known as the interquartile range
5. The half of the interquartile range is quartile deviation. It can also be mentioned as a semi-interquartile range..
6. For calculating the coefficient of quartile deviation, the following formula is applied:
$$Q_3 - Q_1 / Q_3 + Q_1$$

Short Answers:

1. The average of the deviations of all the principles taken from some average value (mean, median, mode) of the series by ignoring the signs (+ or -) of the deviation is known as the mean deviation.
2. It is the square root of the arithmetic mean of the squared deviations of the items from their mean value.
3. A Lorenz curve is a curve that shows the actual distribution deviation (of income or wealth) from the line exhibiting equal distribution.
4. Variance is another measure of dispersion. It is the square of the standard deviation.

Stay tuned for questions papers, sample papers, syllabus, and relevant notifications on our website.

5. Dispersion is the extent to which values in a distribution differ from the average of the distribution. Knowledge of only average is insufficient as it does not reflect the quantum of variation in values.

Measures of dispersion enhance the understanding of a distribution considerably by providing information about how much the actual value of items in a series deviate from the central value, e.g., per capita income gives only the average income but a measure of dispersion can tell you about income inequalities, thereby improving the understanding of the relative living standards of different sections of the society. Through value of dispersion one can better understand the distribution.

Thus a measure of dispersion is a good supplement to the central value in understanding a frequency distribution.

6. Standard Deviation is considered to be the best measure of dispersion and is therefore the most widely used measure of dispersion.

- It is based on all values and thus provides information about the complete series. Because of this reason, a change in even one value affects the value of standard deviation.
- It is independent of origin but not of scale.
- It is useful in advanced statistical calculations like comparison of variability in two data sets.
- It can be used in testing of hypothesis.
- It is capable of further algebraic treatment.

7. Yes, it is true that some measures of dispersion depend upon the spread of values, whereas some calculate the variation of values from the central value. Range and Quartile Deviation measure the dispersion by calculating the spread within which the value lie. Mean Deviation and Standard Deviation calculate the extent to which the values differ from the average or the central value.
8. 25% of the persons earned more than ₹ 45,000. This implies that upper quartile $Q_3 = 45,000$ 75% earned more than 18,000. This implies that lower quartile $Q_1 = 18,000$



Absolute Measure of Dispersion = $Q_3 - Q_1 = 45,000 - 18,000 = 27,000$

Relative Measure of Dispersion

Co-efficient of Quartile Deviation

$$= \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Where, $Q_3 = 3^{\text{rd}}$ Quartile, $Q_1 = 1^{\text{st}}$ Quartile

$$= \frac{45,000 - 18,000}{45,000 + 18,000}$$

$$= \frac{27,000}{63,000} = 0.428$$





7

Correlation

Important terms and concepts

Correlation refers to the associations between variables. When an association exists between two variables, it means that the average value of one variable changes as there is a change in the value of the other variable.

Kinds of correlation –

- ❖ Positive and Negative correlation.
- ❖ Linear and non – linear correlation.
- ❖ Simple and multiple correlations.

Types of Correlation –

- ❖ **Positive correlation:** If two variables change in the same direction (i.e. if one increases the other also increases, or if one decreases, the other also decreases), then this is called a positive correlation. For example: Advertising and sales.

Some other examples of series of positive correlation are:

- Heights and weights.
- Household income and expenditure.
- Price and supply of commodities.
- Amount of rainfall and yield of crops.

- ❖ **Negative correlation:** If two variables change in the opposite direction (i.e. if one increases, the other decreases and vice versa), then the correlation is called a negative correlation. For example: T.V. registrations and cinema attendance.
- ❖ **Linear Correlation:** When two variables change in a constant proportion.
- ❖ **Non- linear correlation:** When two variables do not change in the same proportion.
- ❖ **Simple correlation:** Relationship between two variables are studied.
- ❖ **Multiple Correction:** Relationship between three or more than three variables are studied.

Degrees of Correlation –

- ❖ **Perfect Correlation:** When values of both variables changes at a constant rate.
Types:
 - **Perfect positive correlation:** When values of both variables changes at a constant ratio in the same direction correlation coefficient value (r) is +1.
 - **Perfect negative correlation:** When values of both the variables change at a constant ratio in opposite direction. Value of coefficient of correlation is -1.
- ❖ **Absence of correlation:** If two series of two variables exhibit no relations between them or change in one variable does not lead to a change in the other variable, then we can firmly say that there is no correlation or absurd correlation between the two variables. In such a case the coefficient of correlation is 0. 3. Limited degrees of correlation.

- ❖ **Limited degree correlation:** If two variables are not perfectly correlated or there is a perfect absence of correlation, then we term the correlation as Limited correlation.

Types:

- High: r lies between ± 0.7 & 0.999
- Moderate = r lies between ± 0.5 and $+ 0.699$
- Low: $r < \pm 0.5$

Different methods of finding correlation –

1. Karl Pearson's coefficient method.
2. Rank method/ Spearman's coefficient method.
3. Scatter Diagram.

- ❖ **Karl Pearson's Method:** It gives the precise numerical expression for the measure of correlation. It is denoted by 'r'. The value of 'r' gives the magnitude of correlation and its sign denotes its direction.

Merits of Karl Pearson's Method:

- Helps to find direction of correlation.
- Most widely used method.

Demerits of Karl Pearson's method:

- Based on large number of assumptions.
- Affected by extreme values.

- ❖ **Spearman's Rank Correlation Method:** This method is based on the ranks of the items rather than on their actual values. The advantage of this method over the others is that it can be used even when the actual values of items are unknown. For example if you want to know the correlation between honesty and wisdom of the boys of your class, you can use this method by giving ranks to the boys. It can also be used to find the degree of agreements between the judgments of two examiners or two judges.

Merits of Spearman's Rank Correlation:

- Simple and easy to calculate.
- Not affected by extreme values.

Demerits of Spearman's Rank Correlation:

- Not Suitable for grouped data.
- Not based on original values of observations.

- ❖ **Scatter Diagram:** Scatter Plots (also called scatter diagrams) are used to graphically investigate the possible relationship between two variables without calculating any numerical value. In this method, the values of the two variables are plotted on a graph paper. One is taken along the horizontal (X-axis) and the other along the vertical (Y-axis). By plotting the data, we get points (dots) on the graph which are generally scattered and hence the name 'Scatter Plot'.

Merits of Scatter Diagram:

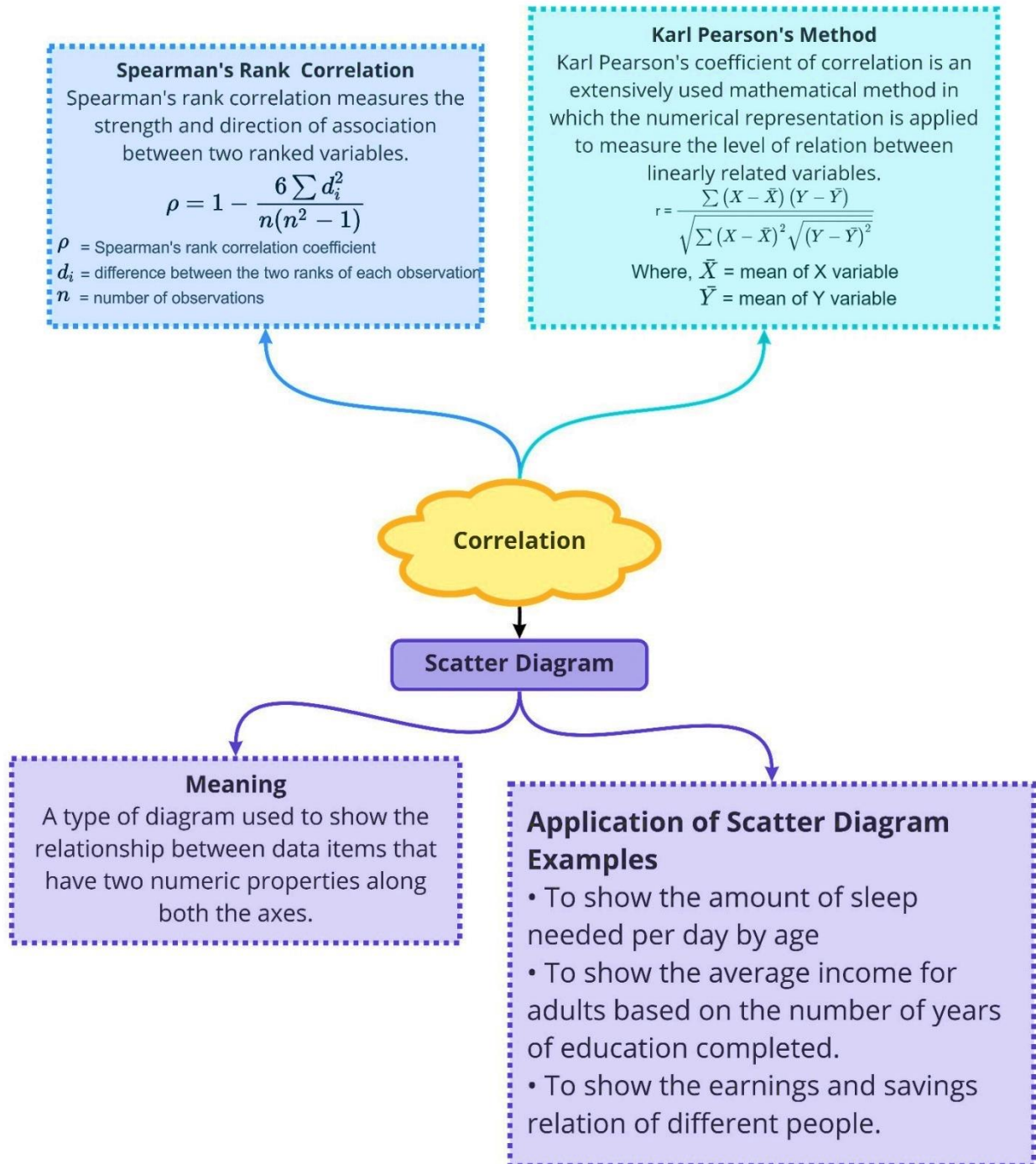
- Most simplest method.
- Not affected by size of extreme values.

Demerits of Scatter Diagram:

- It gives only an approximate idea of the relationship.
- It is only a qualitative expression of the quantitative change.



Class : 11th Economics (Statistics)
Chapter-7 : Correlation



Important Questions

Multiple Choice Questions-

- When the relation of three or more variables is studied simultaneously, it is called:
 - simple correlation
 - partial correlation
 - multiple correlation
 - none of these
- When coefficient of correlation lies between +0.25 and + 0.75, it is called:
 - perfect degree of correlation
 - high degree of correlation
 - moderate degree of correlation
 - low degree of correlation
- Coefficient of correlation lies always between:
 - 0 and +1
 - 1 and 0
 - 1 and +1
 - none of these
- Rank correlation is a superior method of analysis in case of ___ distribution.
 - qualitative
 - quantitative
 - frequency
 - none of these
- When two variables change in a constant proportion, it is called:
 - linear correlation
 - non-linear correlation
 - partial correlation
 - none of these
- A scatter diagram:
 - Is a statistical test
 - Must be linear
 - Must be curvilinear
 - Is a graph of x and y values
- Maximum value of rank correlation coefficient is:
 - 0
 - +1
 - 1
 - None of these
- The correlation coefficient will be -1 if the slope of the straight line in a scatter diagram is:
 - Positive
 - Negative
 - Zero
 - None of these
- In a 'negative' relationship
 - As x increases, y increases
 - As x decreases, y decreases
 - As x increases, y decreases
 - Both (a) and (b)
- Scatter diagram helps us to:
 - Find the nature of correlation between two variables
 - Obtain the mathematical relationship between two variables
 - Compute the extent of correlation between two variables
 - Both (a) and (c)
- The lowest strength of association is reflected by which of the following correlation coefficients?
 - 0.95
 - 0.60
 - 0.35
 - 0.29
- Relation between price and demand is:
 - positive
 - negative
 - one to one
 - no relationship
- If the relationship between x and y is positive, as variable y decreases, variable x:
 - Increases
 - Decreases
 - Remains same
 - Changes linearly
- Kari Pearson's coefficient is defined from:
 - Ungrouped data
 - Grouped data
 - Both (a) and (b)
 - None of these
- When $r = 1$, all the points in a scatter diagram would lie:
 - On a straight line directed from lower left to upper right
 - On a straight line
 - On a straight line directed from upper left to lower right
 - Both (a) and (b)



Very Short Questions:

1. Define correlation.
2. Define partial correlation.
3. Define the line of best fit.
4. Explain the principal methods for calculating the coefficient of correlation?
5. What is the difference between negative and positive correlation?
6. What is the nature of correlation of two variables, when they move in the same direction?
7. Coefficient of correlation is between -1 and +. How would you express it arithmetically?
8. When is the rank correlation method used?
9. What is a simple correlation?
10. What are the multiple correlations?

Short & Long Questions:

1. Concept and definition of correlation?
2. What does Correlation Measure?
3. Linear Correlation?
4. Non-Linear Correlation?
5. Degree of correlation?
6. Importance or significance of correlation?

Assertion Reason Question:

1. Read the following statements given below and choose the correct alternative.

- (a) Both assertion and reason are true. The reason is the correct explanation of the assertion
- (b) Both assertion and reason are not true. The reason is not the correct explanation of the assertion
- (c) Assertion is true but the reason is not
- (d) Reason is true but the assertion is not

Assertion- Quantitative method of calculating correlation is given by Karl Pearson.

Reason- Karl Pearson's coefficient of correlation is generally written as 'y'.

2. Read the following statements given below and choose the correct alternative.

- (a) Both assertion and reason are true. The reason is the correct explanation of the assertion
- (b) Both assertion and reason are not true. The reason is not the correct explanation of the assertion
- (c) Assertion is true but the reason is not
- (d) Reason is true but the assertion is not

Assertion- Karl Pearson formula applies only to those series that comes out from the actual average.

Reason- the value of the coefficient of correlation can vary between + 1 and -1

STEP UP
ANSWER KEY
ACADEMY

Multiple Choice Answers-

1. C
2. C
3. C
4. A
5. A
6. D
7. B
8. B
9. C
10. A
11. D
12. B
13. B
14. A
15. A

Very Short Answers:

1. Correlation is a statistical method or a technique that measures, a quantitative relationship between different variables, such as demand and price.
2. When more than two variables are involved, and out of these, the relationship between only two variables is suited to treating other variables as constant, then the correlation is partial.
3. The line of the best fit is one of that passes through the scattered points such that it represents most of these points. Roughly, half of the scattered points should be on either side.
4. The principal methods for calculating the coefficient of correlation are.
 - Scattered Diagram Method
 - Karl Person's Coefficient of Correlation, and
 - Spearman's Rank Correlation Coefficient.

5. The difference between negative and positive correlation is that in positive correlation, variables move in the same direction, whereas, in the negative correlation, they move in the opposite direction.
6. When two variables move in the same direction, such correlation is positive.
7. When the coefficient of correlation is -1, correlation is perfect negative, and when the coefficient of correlation is +1, correlation is perfect positive.
8. The rank correlation method is used when variables are qualitative, such as bravery, beauty, virtue, and wisdom, etc.
9. A simple correlation implies the study or a relationship between two variables only.
10. When the relation between three or more variables are studied simultaneously, it is known as multiple correlations.

Short & Long Answers:

1. **CONCEPT AND DEFINITION OF CORRELATION**

The statistical methods so far studied in this book focus on the analysis of one variable or one statistical series only. In real life however, two or more than two statistical series may be found to be mutually related. For instance, change in price leads to change in quantity demanded. Increase in supply of money causes increase in price level. Increase in level of employment results in increase in output. Such situations necessitate simultaneous study of two or more statistical series. The focus of study in such situations is on the degree of relationship between different statistical series. The statistical technique that studies the degree of such relationships is called the technique of correlation.

Definition

According to Croxton and Cowden, "When the relationship is of a quantitative nature, the appropriate statistical tool for discovering and measuring the relationship and expressing it in a brief formula is known as correlation." In the words of Boddington, "Whenever some definite connection exists between the two or more groups, classes or series or data there is said to be correlation."

Relationship between Two Variables may just be a Coincidence One may find a relationship between two variables which is just a coincidence. Example: When there is a departure of migratory

birds from a sanctuary, you may find a fall in wedding ceremonies in the country. Such relationships are meaningless. These are in other words, spurious relationships which are devoid of any meaningful conclusion. Such relationships are not to be treated as correlations, Only those relationships are to be treated as correlations which offer some meaningful conclusions.

Example: Increase in rainfall and increase in rice production is a relationship that makes sense: increase in per capita income and decrease in death rate is a meaningful relationship; Good percentage of marks in physics may be related to good percentage of marks in mathematics; and so on.

Positive and Negative Correlation Correlation between different variables may either be positive or negative. Here is a brief description of the two:

(1) Positive Correlation : When two variables move in the same direction, that is, when one increases the other also increases and when one decreases the other also decreases, such a relation is called positive correlation.

Relationship between price and supply may be cited as an example. Check the following table as an illustration:

(A) Simultaneous increase in the values of both variables		(B) Simultaneous decrease in the values of both variables	
X	Y	X	Y
10	100	50	200
20	150	40	150
30	200	30	100
40	250	20	50

(2) Negative Correlation : When two variables change in different directions, it is called negative correlation. Relationship between price and demand, may be cited as an example. The following table demonstrates this relation:

Rise in the Value of One Variable is X	Accompanied with a fall in the other Y
1	5
2	4
4	2
5	1



2. Often the students tend to believe that correlation suggests a relationship between two variables where one is the cause of the other. Example: This is correlation between price and quantity demanded of a commodity. Clearly, an increase in price causes a decrease in quantity demanded, and vice versa. Change in price causes changes in quantity demanded. But to be emphatically noted is the point that cause and effect relationship between the variables is not at all any pre-condition in the theory of correlation.

Correlation just measures the degree and intensity of relationship between the two variables, with or without any cause and effect relationship. Of course, the established relationship between the variables should be capable of offering us some meaningful conclusion. Example: Students who are good in academics may be good in sports also. Certainly, it is a meaningful relationship (or correlation) if one finds it. But surely there is no cause and effect relationship between the two variables. Linear and Non-Linear Correlation.

3. Linear Correlation

When two variables change in a constant proportion, it is called linear correlation. If the two sets of data bearing fixed proportion to each other are shown on a graph paper, their relationship will be indicated by a straight line. Thus, linear correlation implies a straight-line relationship.

Illustration.

Linear Correlation

(a)	2	4	6	8	10	12	14
(b)	5	10	15	20	25	30	35
(a)	2	4	6	8	10	12	14
(b)	3	7	12	18	25	35	45

Thus, for every change in variable (a) by 2 units there is a change in variable (b) by 5 units.

4. Non-linear Correlation

Here there is no specific relationship between the two variables, though both tend to change in the same direction. That is, both are increasing, but not in any constant proportion.

Simple and Multiple Correlation

(1) Simple Correlation : Simple correlation implies the study of relationship between two variables only. Like the relationship between price and demand or the relationship between money supply and price level.

(2) Multiple Correlation : When the relationship among three or more than three variables is studied simultaneously, it is called multiple correlations. In case of such correlation, the entire set of independent and dependent variables is simultaneously studied. For instance, effects of rainfall, manure, water, etc., on per hectare productivity of wheat are simultaneously studied.

5. Degree of correlation refers to the Coefficient of Correlation. There can be the following degrees of positive and negative correlation.

(1) Perfect Correlation: When two variables change in the same proportion it is called perfect correlation. It may be of two kinds:

(i) Perfect Positive : Correlation is perfectly positive when proportional change in two variables is in the same direction. In this case, coefficient of correlation is positive (+1).

(ii) Perfect Negative: Correlation is perfectly negative when proportional change in two variables is in the opposite direction. In this case, coefficient of correlation is negative (-1).

(2) Absence of Correlation: If there is no relation between two series or variables, that is, change in one has no effect on the change in other, then those series or variables lack any correlation between them.

(3) Limited Degree of Correlation: Between perfect correlation and absence of correlation there is a situation of limited degree of correlation. In real life, one mostly finds limited degree of correlation. Its coefficient (r) is more than zero and less than one ($r > 0$ but < 1). The degree of correlation between 0 and 1 may be rated as:

(i) High: When correlation of two series is close to one, it is called high degree of correlation. Its coefficient lies between 0.75 and 1,

(ii) Moderate: When correlation of two series is neither large nor small, it is called moderate degree of correlation. Its coefficient lies between 0.25 and 0.75.

(iii) Low: When the degree of correlation of two series is very small, it is called low

degree of correlation. Its coefficient lies between 0 and 0.25. All these degrees of correlations may be positive or negative.

Degree of Correlation

Degree	Positive	Negative
Perfect	+1	-1
High	Between +0.75 and +1	Between -0.75 and -1
Moderate	Between +0.25 and +0.75	Between -0.25 and -0.75
Low	Between 0 and +0.25	Between 0 and -0.25
Zero	0	0

6. Following observations highlight the importance or significance of correlation as a statistical method:

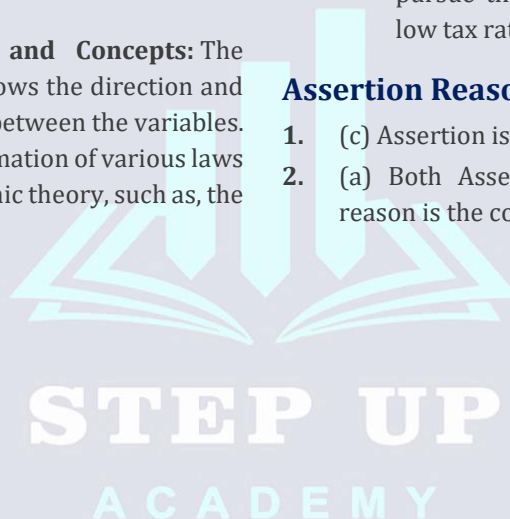
(1) Formation of Laws and Concepts: The study of correlation shows the direction and degree of relationship between the variables. This has helped the formation of various laws and concepts in economic theory, such as, the

law of demand and the concept of elasticity of demand.

- (2) Cause and Effect Relationship:** Correlation coefficient sometimes suggests cause and effect relationship between different variables. This helps in understanding why certain variables behave the way they behave.
- (3) Business Decisions:** Correlation analysis facilitates business decisions because the trend path of one variable may suggest the expected changes in the other. Accordingly, the businessman may plan his business decisions for the future.
- (4) Policy Formulation:** Correlation analysis also helps policy formulation. If the Government finds a negative correlation between tax rate and tax collection, it should pursue the policy of low tax rate. Because, low tax rate would lead to high tax collection.

Assertion Reason Answer:

- (c) Assertion is true but the reason is not.
- (a) Both Assertion and Reason are true. The reason is the correct explanation of the assertion.





8

Index Numbers

Introduction to Index Number

An index number is a statistical device for measuring changes in the magnitude of a group of related variables.

Characteristics of Index Numbers –

- ❖ Index numbers are not qualitative statements like prices are rising or falling. It is a precise measurement of quantitative changes in the concerned variable.
- ❖ Index numbers show changes in terms of averages. For example when it is said that price level has been increased it does not mean that price of all goods and services have been increased. But it means that on and average prices have been increased.
- ❖ An Index number, indicating change in magnitude, as of price, wage, employment, or production shifts, relative to the magnitude at a standard or base value usually taken as 100.

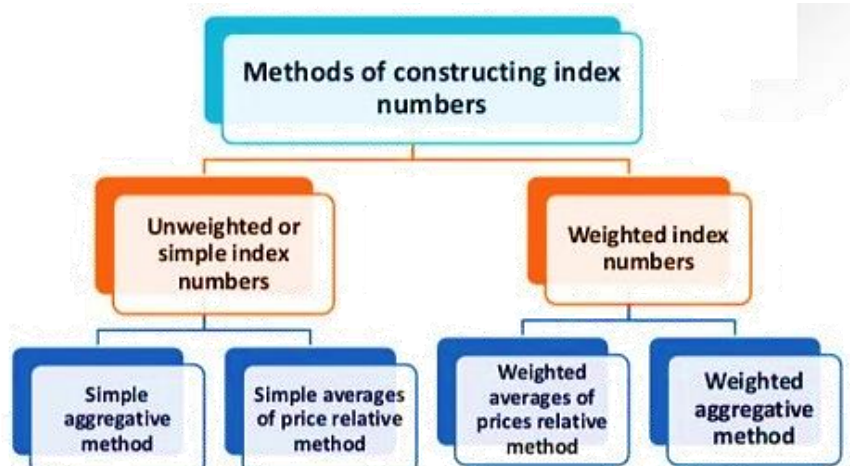
Types of Index Number –

Index numbers are named after the activity they measure.

Their types are as under:

- ❖ **Price Index:** Measure changes in price over a specified period of time. It is basically the ratio of the price of a certain number of commodities at the present year as against base year. Some price index numbers are Wholesale price Index (WPI), Consumer Price Index (CPI) or Cost of Living Index (COLI).
- ❖ **Quantity Index:** As the name suggests, these indices pertain to measuring change in volume of commodities like goods produced or goods consumed, etc. An important quantity index number is Index of Industrial Production (IIP).
- ❖ **Value Index:** These pertain to compare changes in the monetary value of imports, exports, production or consumption of commodities.

Methods of Constructing Index Numbers –



Weighted Aggregative Method -

(i) Laspeyre's Method : $P_{01} = \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100$

(ii) Pasche's Method : $P_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100$

(iii) Fisher's Method : $P_{01} = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times 100$

Fisher's Index Number - Is called ideal index number because:

- It considers both base year and current year quantity;
- It is based on Geometric Mean which is considered as best average.
- It satisfies time reversal & factor reversal test.

Weighted Average of Price Relative -

$$P_{01} = \frac{\sum RW}{\sum W}$$

Where $R = \frac{P_1}{P_0} \times 100$

W = Weight

❖ **Wholesales Price Index (WPI):**

WPI is used to measure the relative changes in the prices of commodities traded in the wholesale markets. presently 2011-12 is used as the base year. It is also used to calculate the rate of inflation in a country.

❖ **Consumer Price Index (CPI) OR Cost of living Index (COLI):**

CPI can be measured through two methods:

- Aggregate expenditure method.
- Family budget method

$$CPI = \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100$$

Family budget method:

$$CPI = \frac{\sum RW}{\sum W}$$

If W is not given ($W = p_0 q_0$)

$$R = \frac{p_1}{p_0} \times 100$$

❖ **Index of Industrial Production (IIP):**

The index of industrial production is a composite indicator that measures the short-term changes in the quantity of production of industrial products during a given period with respect to that in a chosen base period.

$$IIP = \frac{\sum \left(\frac{Q_1}{Q_0} \right) W}{\sum W} \times 100$$



Where Q_1 = Production level in current year
 Q_0 = Production level in base year
 W = Weightage of different industrial output

❖ **Inflation and Index Number:**

Inflation is the percentage increases in price level i.e. prices of basket of goods and services over a specific period of time i respect of base year.

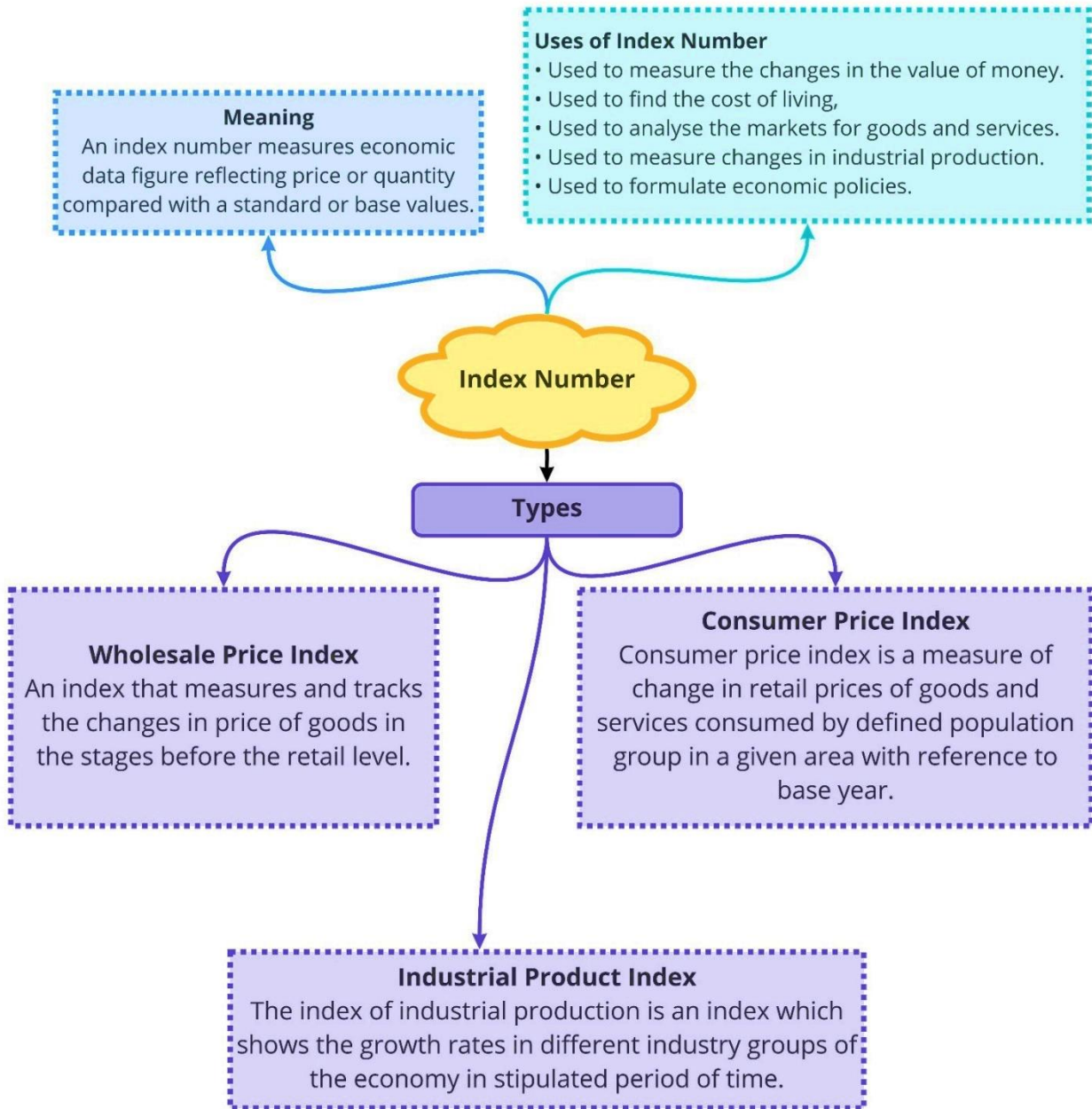
$$\text{Inflation Rate} = \frac{l_1 - l_0}{l_0} \times 100$$

Where, l_1 = Index of current period

l_0 = Index of base period



Class : 11th Economics (Statistics)
Chapter-8 : Index Number



15. Cost of living index is known as
- cost price index
 - consumer price index
 - both
 - none

Very Short Questions:

- Define index number.
- What is a simple index number?
- Define the weighted index number.
- Explain price relative.
- Define consumer price index number.
- What is the wholesale price index?
- State the two types of price index numbers.
- What should be the base year like?
- Name the consumer groups for which consumer price index number is computed.
- Name one principal limitation of index numbers.

Short & Long Questions:

- Concept and definition of index numbers?
- Features or characteristics of index numbers?
- Difficulties or problems in the construction of index numbers?
- Advantages or uses of index numbers?
- limitations of index numbers?
- Simple and weighted index numbers?
- Methods of constructing index numbers?
- Importance of the Consumer Price Index or Cost of Living Index?
- Difficulties in the Construction of Consumer Price Index?

11. Wholesale Price Index (WPI)

Assertion Reason Question:

- Read the following statements given below and choose the correct alternative.
 - Both assertion and reason are true. The reason is the correct explanation of the assertion
 - Both assertion and reason are not true. The reason is not the correct explanation of the assertion
 - Assertion is true but the reason is not
 - Reason is true but the assertion is not

Assertion- wholesale price index measures the relative changes in the price of commodities traded in the wholesale markets.

Reason- wholesale price index is used for forecasting demand and supply.

- Read the following statements given below and choose the correct alternative.

- Both assertion and reason are true. The reason is the correct explanation of the assertion
- Both assertion and reason are not true. The reason is not the correct explanation of the assertion
- Assertion is true but the reason is not
- Reason is true but the assertion is not

Assertion- Inflation refers to the situation of rising the general price level in our country over a fairly long period of time.

Reason- Inflation is measured in terms of a consumer price index.

ANSWER KEY

Multiple Choice Answers-

- C
- C
- D
- A
- A
- C
- D
- B
- B
- A
- A
- D
- A
- A
- B

Very Short Answers:

- An index number is a mathematical measure outlined to explain changes in a group of related variables or just variable considering the time, characteristics, and geographical location.



2. Simple index numbers are the index numbers in which all items of the series are accorded equal weightage or importance.
3. These are the index number in which different item of the series are accorded different weightage, depending upon their relative.
4. A price relative is the percentage ration of the value of a variable in the current year to its value in the base year.
5. It is the index number which measures the average change in the prices. The specific class of consumer pays this price for goods and services consumed by then in the current year in comparison with a base year.
6. The wholesale price index calculates the relative difference in the prices of goods traded in the wholesale markets.
7. The two types of price index numbers are.
 - Consumer Price Index
 - Wholesale Price Index
8. The base year should be a year without wide fluctuations, neither very long nor concise period of study and for which reliable data are available.
9. The consumer groups for which consumer price index number is computed are.
 - Industrial Workers
 - Urban-Non Manual Employees
 - Agricultural Labourers
10. One principal limitation of index numbers is that it owes a difference in unit of currency and the composition of production across the world. It is difficult to construct an index number that facilitates international comparison.

Short & Long Answers:

1. The concept of index number can be best understood through an illustration. Let us consider a situation of rising prices during the year 2018. In this context, we are faced with three basic questions. First, compared to which year have the prices risen during 2018? Second, how do we handle the situation when the prices of some goods rise more than the others? Thirdly, can prices of different goods be expressed in terms of any standard unit or different units are to be used to express prices of different goods and services, such that the price of milk is to be expressed in terms of rupees per liter, of cloth in terms of rupees per meter and of sweets in terms of rupees per kilogram. The study of Index Numbers

answers all these questions: First, rise in prices during 2018 would be studied only with reference to some previous years like 2001 or 2004. Otherwise, the mere statement that prices during 2018 have tended to rise would make no sense. 2018 will be treated as the current year and 2001 or 2004 as the base year. Prices during the base year are taken as 100. Prices during the current year are related to the base year price. So that, we find out percentage change in price level over the years. Level of price is called the index of price. Since price of the base year is assumed to be 100, we can say that index of price of the base year is always 100. If base year is 2004 and the price index is 100, and if in the year 2018 prices are doubled, we shall conclude that the index of price in the year 2018 has shot up to 200 compared to 100 in the base year. As regards the second question (how do we handle the situation when prices of some goods change more than the others) the study of index numbers suggests to take average change. Thus, if the price of Potatoes has rises from 100 to 200 and the price of Onions has risen from 100 to 300, we shall consider average change only, viz. $200+300/2 = 250$

Thus, it is the average index of prices for the various goods and services which is constructed for different years, and compared with the base year.

Third, as regards the problem of expressing the prices of various goods and services in some homogeneous units, the theory of Index Numbers suggests to consider only percentage change in prices of various goods and services. Once a change in price is expressed as a percentage change, the unit of the commodity (like litre of a milk, or meter of a cloth) loses its relevance.

Thus, what does the study of Index Numbers do? It helps us to find out percentage change in the values of different variables (may be prices of different goods or production of different commodities) over time with reference to some base year which happens to be the year of comparison. When various goods are studied simultaneously, the percentage change is taken as the average for all the goods.

Definition

In the words of Spiegel, "An Index Number is a statistical measure designed to show changes in a variable or group of related variables with respect to time, geographic location or other characteristics".

According to Croxton and Cowden, "Index Numbers are devices for measuring difference in the magnitude of a group of related variables".

2. Following are the three specific features or characteristics of index numbers:

(1) **Relative Changes:** Index numbers measure relative or percentage changes in the variable(s) over time. Index number of prices, for example, is not simply a statement of prices at different dates, it presents estimates of percentage changes in the prices over years with reference to some selected base year. If index of prices stands at 200 in 2018 compared to 100 in 2004-05 (the base year), it suggests that compared to the base year, prices have risen by 100 per cent.

(2) **Quantitative Expression:** Index numbers offer a precise measurement of the quantitative change in the concerned variable(s) over time. The index of prices, for example, will tell us that between the years 2017 and 2018, prices have risen by 7 percent, or that industrial production has declined by 3 per cent or that national income has risen just by 3.5 per cent during this period. The index numbers are not the qualitative statements like prices are rising or falling.

(3) **Averages:** Index numbers show changes in terms of averages. For example, when it is said that between the years 2017 and 2018, prices have risen by 7 per cent, it does not mean that prices of all goods and services have uniformly risen by 7 per cent; it only means that on an average there has been a 7 per cent rise in the prices of various goods and services. Even when prices of certain goods might have risen by more than 7 per cent and of certain others by less than 7 per cent.

3. DIFFICULTIES OR PROBLEMS IN THE CONSTRUCTION OF INDEX NUMBERS

(1) **Purpose of Index Number:** There are various types of index numbers, constructed with different objectives. Before constructing an index number, one must define the objective. The construction of index number is significantly influenced by the objective or purpose of the study. Thus, for example, if the objective is to study the impact of change in the value of money on the consumers one should construct consumers' price index

number. If we are to study the impact of change in the purchasing power of money on the producers, we shall construct index number on the basis of wholesale prices. Haberler has rightly pointed out that, "Different index numbers are constructed to fulfil different objective and before setting to construct a particular index number, one must clearly define one's object of study because, it is on the objective of the study, that the nature and format of the index number depends."

(2) **Selection of Base Year:** Selection of Base Year is another problem in the construction of index number. Base year is the reference year. It is the year with which prices of the current year are compared. As far as possible, Base Year should be a normal year. That is, it should be the one without much ups and downs. Otherwise, the index values would fail to capture the real change in the variable. The year 2004-05 is treated as base year in India, at present.

(3) **Selection of Goods and Services:** Having defined the objective, the problem is of the selection of goods or Services to be included in the index number. To construct the Consumers' Price Index, for example, all commodities are not included. It is neither possible nor desirable to include all the goods and services produced in the country. We have to choose those goods and services which represent most of

Purpose of Constructing Index Number of

(i) Prices, and (ii) Quantities

(i) Purpose of constructing index number of prices is to know the relative change or percentage change in the price level (made up of simple or weighted average of the prices of different goods and services) over time.

A rising general price level over time is a pointer towards inflation, while a falling general price level is a pointer towards deflation. Both inflation as well as deflation have notable consequences for an overall economic activity in the economy.

(ii) Purpose of constructing index number of quantity is to know relative change or percentage change in the quantum or volume



of output of different goods and services over time. This reflects the level of economic activity in the economy and its different sectors. A rising index of quantity suggest a rising level of economic activity and vice versa others in the market. In other words, commodities selected should be such as are widely consumed, for example, rice, milk, ghee, cloth, etc. Larger the number of goods and services more representative is the index number.

- (4) **Selection of the Prices of the Goods and Services:** Having selected the goods and services, the problem arises of prices to be selected. Broadly, in the construction of Price Index, the problem is whether to adopt retail prices or wholesale prices, controlled or open market prices. The choice would depend upon the objective or purpose of the study.
- (5) **Finding the Average Prices:** In the construction of index number, base year value is assumed to be 100 and other values of different years are related to 100. Thus, if cloth price is Rs. 5 per meter in the base year and is found to be Rs. 10 per meter in the current year, the index of prices of cloth would be $10/5 \times 100 = 200$ for the current year.

Likewise, price relatives for other commodities are worked out and average for these price relatives is determined and compared with the base year value of 100. It may be noted here that average of base year remains 100, but the average of the year under investigation may be more or less than 100. In case the average of the year under investigation is more than the average of the base year, it means that general price level has gone up. If it is less than the base year, it means that general price level has gone down. Generally, base year is indicated as '0' and current year as T\ Price index is written as P01 and it is read as price index of year 1 in relation year 0.

- (6) **Selection of Weights:** While constructing index number, weights are accorded to different commodities according to their relative significance. There are several methods of according weight, e.g., Fisher's method, Paasche's method, Laspeyre's method. While constructing weighted index number, one must justify his choice of

weighting technique in accordance with the nature and objective of his study.

- (7) **Choice of Average:** In finding out average values, different kinds of average may be used, geometric average, arithmetic average, etc. The choice of average significantly influences the results. Different kinds of averages may give different index number of a given change in price.
- (8) **Selection of Formula:** Index numbers can be constructed with the help of many formulae, such as, Laspeyre's method, Paasche's method, Dorbish and Rowley's method, Fisher's method. One has to decide about the method to be used while constructing the index number.

4. ADVANTAGES OR USES OF INDEX NUMBERS

Some of the main advantages or uses of index numbers are as under.

- (1) **Measurement of Change in the Price Level or the Value of Money:** Most important use of index numbers is that index numbers measure the value of money during different periods of time. We can use index numbers to know the impact of the change in the value of money on different sections of the society. Accordingly, devices or means can be worked out to correct inflationary or deflationary gaps in the system.
- (2) **Knowledge of the Change in Standard of Living:** Index numbers help to ascertain the living standards of people. Money incomes may increase but if index numbers show a decrease in the value of money, living standards may even decline. Thus, index numbers indicate change in real income.
- (3) **Adjustments in Salaries and Allowances:** Cost of living index is a useful guide to the Government and Private Enterprises to make necessary adjustments in salaries and allowances of the workers. Increase in the cost of living index suggests increase in salaries and allowances.
- (4) **Useful to Business Community:** Price index numbers serve as a useful guide to the business community in their planning and decisions. Trend of the prices significantly influence their production decisions.
- (5) **Information Regarding Production:** Index numbers of production shows whether the level of agricultural and industrial

production in the economy is increasing or decreasing. Accordingly, agricultural and industrial development policies are formulated.

- (6) **Information Regarding Foreign Trade:** Index of exports and imports provides useful information regarding foreign trade. Accordingly, export-import policies are formulated.
- (7) **Useful to Politicians:** Politicians come to know of the real economic condition in the country on the basis of index numbers. They offer constructive criticism of government's economic policies and give suggestions for economic reforms in the country.

5. LIMITATIONS OF INDEX NUMBERS

In the construction of index numbers, there are some practical difficulties and theoretical limitations. The same are as under:

- (1) **Not Completely True:** Index numbers are not fully true. For example, one can only make an estimate of change in the value of money with the help of index numbers. The index numbers simply indicate arithmetical tendency of the temporal changes in the variable.
- (2) **International Comparison not Possible:** Different countries have different basis of index numbers. These do not help international comparisons.
- (3) **Difference of Time:** With the passage of time, it is difficult to make comparisons of index numbers. With the changing times, man's habits, tastes, etc., also undergo a change. Consequently, index numbers constructed on the basis of old consumption pattern cannot be compared with the index numbers constructed on the basis of new consumption pattern.
- (4) **Limited Use:** Index numbers are prepared with certain specific objective. If they are used for another purpose they may lead to wrong conclusions. For example, index numbers prepared to know about the economic condition of the teachers cannot be used to know about the economic condition of the labourers.
- (5) **Lack of Retail Price Index Numbers:** Most of the index numbers are prepared on the basis of wholesale prices. But in real life, retail prices are most relevant, but it is

difficult to collect retail prices. Index numbers based on wholesale prices may be misleading. With regard to the limitations of index numbers, Coulbourn has rightly said, "In this changing world it is difficult to escape from the theoretical defects and in future, as far as we can see, it will not be possible, from theoretical point of view, to make use of the best method, of constructing the index number."

6. SIMPLE AND WEIGHTED INDEX NUMBERS

'Simple' and 'weighted' are the two broad categories of index numbers. Here is a brief description of these concepts.

Simple Index Numbers

These are the index numbers in which all items of the series are accorded equal weightage or importance. In case of a simple index of prices, for example, all goods and services are to be accorded equal weightage, no matter whether sale/purchase of certain goods is more than that of the others. It will be a simple average of the prices of different goods and services.

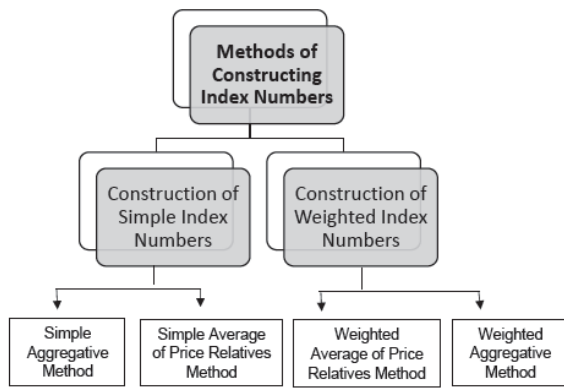
Weighted index Numbers

These are the index numbers in which different items of the series are accorded different weightage, depending upon their relative importance. It is not a simple average of prices of different goods and services, as in case of a simple price index. Instead, it is to be a weighted average of the prices of different goods. Thus, if the expenditure on rice is twice the expenditure on cloth, then in the construction of price index, price of rice may be accorded '2' as the weight compared to the weightage of '1' to the price of cloth. Though difficult to construct, weighted index numbers certainly offer a much more realistic view of the change over time compared to the simple index numbers.

The basic difference between Simple Index and Weighted Index In the simple index, all items of the series are treated as of equal importance. In the weighted index, weights are accorded to different items depending on their relative importance.

7. METHODS OF CONSTRUCTING INDEX NUMBERS

The following chart shows the various methods of constructing index numbers (Simple as well as weighted):



Let us attempt a brief description of the various methods.

Base Year and its Characteristics : Base year is the year of comparison, also called reference year. It should bear the following characteristics:

- (i) It should be a normal year, not showing wide fluctuations in the parameters related to the index number.
- (ii) It should be a year for which reliable statistical data are available, so that comparison of the performance of the other years with the base year becomes meaningful-
- (iii) It should not be a year too far from the period of study. Otherwise, relative change over time would not make much sense,
- (iv) It should be neither very long nor very short period. Generally, it is not more than a year and not less than a month.

CONSTRUCTION OF SIMPLE INDEX NUMBERS

There are two methods of constructing simple index numbers:

(1) Simple Aggregative Method : In this method, aggregate of the prices of commodities in the current year are divided by the aggregate of their prices in the base year and multiplied by 100 to get index value for the current year. It is expressed by the following formula:

FORMULA

$$P_{01} = \frac{\sum P_1}{\sum P_0} \times 100$$

Here, P_{01} = Price index of the current year.

$\sum P_1$ = Sum of the prices of the commodities in the current year.

$\sum P_0$ = Sum of the prices of the commodities in the base year.

Current Year: Current year is the year for which average change is to be measured or index number is to be calculated.

Base Year: Base year is the year of reference from which we want to measure extent of change in the current year. The index number of base year is generally assumed to be 100.

Illustration.

Given the following data and assuming 2004 as the base year, find out index value of the prices of different commodities for the year 2018.

Commodity	A	B	C	D	E
Price in 2004 (Rs.)	50	40	10	5	2
Price in 2018 (Rs.)	80	60	20	10	6

Solution:

Construction of a Simple Index Number

Simple Aggregative Method

Commodity	2004 Price (Rs.) (P_0)	2018 Price (Rs.) (P_1)
A	50	80
B	40	60
C	10	20
D	5	10
E	2	6
Total	$\sum P_0 = 107$	$\sum P_1 = 176$

$$P_{01} = \frac{\sum P_1}{\sum P_0} \times 100$$

$$= \frac{176}{107} \times 100$$

$$= 164.49$$

Price Index = 164.49.

(2) Simple Average of Price Relatives Method

According to this method, we first find out price relatives for each commodity and then take simple average of all the price relatives.

What is Price Relative?

A price relative is the percentage ratio of the value of a variable in the current year to its value in the base year. In other words, a price relative is a percentage ratio between price of a commodity in the current year and that in the base year.

Price Relatives, $P_{01} = \frac{\text{Current Year Price } (P_1)}{\text{Base Year Price } (P_0)} \times 100$
We can find out price index number of the current year by using the following formula.

FORMULA

$$P_{01} = \frac{\sum (P_1/P_0 \times 100)}{N}$$

An Important Caution

Prices for different commodities are expressed with reference to different units of measurement. Price of steel, for example, is expressed as rupee per kg, while the price of milk is expressed as rupee per liter, and the price of doth is expressed as rupee per meter. Simple aggregate method of index numbers cannot be used for commodities with different units of measurement. It can be used only for those commodities which have a common unit of measurement.

(Here,
 $P1/P0 \times 100 \times 100 =$ Price relatives; N = Number of goods; P, = Current year's value; P0 = Base year's value.)

Illustration.

Given the following data and using the Price Relatives Method, construct an index number for the year 2018 in relation to 2004 prices.

Commodity	Wheat	Ghee	Milk	Rice	Sugar
2004 Price (Rs.)	100 (per qt.)	8 (per kg.)	2 (per l.)	200 (per qt.)	1 (per kg.)
2018 Price (Rs.)	200	40	16	800	6

Notation of 0 and 1 Often the notation of 0 stands for the base year, and 1 stands for the current year.

Solution:

Construction of a Simple Index Number

Simple Average of Price Relative Method

Commodity	Base year 2004 Price (P ₀)	2018 Price (P ₁)	Price relatives of 2018 in relation to 2004 $\left(\frac{P_1}{P_0} \times 100\right)$
Wheat	100 (per qt.)	200 (per qt.)	$\frac{200}{100} \times 100 = 200$
Ghee	8 (per kg.)	40 (per kg.)	$\frac{40}{8} \times 100 = 500$
Milk	2 (per l.)	16 (per l.)	$\frac{16}{2} \times 100 = 800$
Rice	200 (per qt.)	800 (per qt.)	$\frac{800}{200} \times 100 = 400$
Sugar	1 (per kg.)	6 (per kg.)	$\frac{6}{1} \times 100 = 600$
N = 5			$\Sigma \left(\frac{P_1}{P_0} \times 100\right) = 2,500$

$P01 = \Sigma(P1/P0 \times 100)/N = 2,500/5 = 500$

Price Index = 500.

CONSTRUCTION OF WEIGHTED INDEX NUMBERS

There are two methods of constructing weighted index numbers, as discussed below:

(1) **Weighted Average of Price Relatives Method**

Not all Commodities are Ever Included

Not all commodities are ever included in the construction of an Index Number. Only a sample of commodities is taken which represents characteristics of the entire group of commodities under study.

According to this method, weighted sum of the price relatives is divided by the sum total of the weights. In this method, goods are given weight according to their quantity. Thus,

FORMULA

$P01 = \Sigma RW / \Sigma W$

(Here, P01 = Index number for the current year in relation to the base year; W = Weight; R = Price relative.)

8. Importance of the Consumer Price Index or Cost of Living Index

(1) **Formulation of Price Policy:**

The consumer price indices are used by government to frame policies on prices. On the basis of these indices government decides whether the prices are to be controlled, dual price policy should be adopted or public distribution system is to be introduced, etc. Also, government policies like rent control and taxation, general economic and fiscal policies etc. are framed on the basis of the consumer price index numbers to a large extent.

(2) **Wage Adjustment:**

Cost of living index numbers are used as basis for the wage adjustments. The rates of dearness allowances are decided by the government on the basis of these indices. These indices are also used for wage contracts and wage agreements of the workers.

(3) **Measurement of Real Value:**

These index numbers are used to measure the real value of the rupee or its purchasing power and real income (or revenue), etc.

(4) **Analysis of Markets:**

The consumer price indices are also used for the analysis of the market of specific commodities for their demand and supply.



(5) **National Income Deflator:** These indices are also used as deflators of national income. Accordingly, real change in national income is estimated.

9. Difficulties in the Construction of Consumer Price Index

There are many difficulties in the construction of consumer price index. These are as follows:

- (1) **Difference in the Standards of Living:** Consumption pattern of different classes of consumers are different and therefore, have different living standards. Thus, there cannot be one consumer price index number for different classes of society.
- (2) **Difference in Prices:** The price indices are constructed on the basis of retail prices. But the retail prices vary from place to place and even at the same place from shop to shop. As such it is very difficult to find a representative price for the calculation of consumer price index.
- (3) **Difference in the Proportion of Expenditure:** All the members of any particular group do not spend on various items of consumption in same ratio and even one person does not spend on various commodities in the same ratio at two different periods of time. A consumer's purchase ratio depends upon his/her taste, habits, etc. Accordingly, it is difficult to construct a cost of living index that truly reflects a change in the cost of living over time.

10. WHOLESALE PRICE INDEX (WPI)

The Wholesale Price Index (WPI) measures the relative changes in the prices of commodities traded in the wholesale markets. In India, the wholesale price index numbers are constructed on weekly basis. The year 2011-12 is being used as the base year.

Commodity Group and Weightage of Wholesale Price Index

In India, all the commodities have been classified in the following three groups:

Commodity Group	Name of Commodities	Weightage
(i) Primary Articles	These include 98 commodities like Rice, Fruits, Pulses, Vegetables and Non-food articles like Cotton, Jute, Metals.	22.02

(ii) Fuel, Power, Light and Lubricants	These include 19 items like Coak, Petroleum Products, Electricity, LPG.	14.23
(iii) Manufacturing	It includes 318 items like Textiles, Sugar, Paper, Machinery, Chemicals, Fertilizers, Leather, etc.	63.75

The basic difference of purpose behind the Consumer Price Index Number and the Wholesale Price Index Number

In case of consumer price index number, the basic purpose is to know cost of living of a specified group of consumers in the society. In case of wholesale index number, the basic purpose is to assess the situations of overall demand and supply in the economy.

Rising prices indicate a situation of excess demand, while falling prices suggest a situation of deficient demand. Wholesale price index focuses on the rate of inflation in the economy.

Producer Price Index As in many countries, in India also, efforts are afoot to shift from WPI (Wholesale Price Index) to Producer Price Index. Producer-Price refers to the basic price including taxes, trade margins and transport cost. Producer Price Index is expected to offer better insights into the analysis of price trends in the country.

Uses of Wholesale Price Index

(1) **Forecasting Demand and Supply:** The wholesale price indices are often used to forecast demand and supply situation in the economy. An increase in wholesale price index is an indication of excess demand. It is a situation in which demand is greater than supply. On the other hand, a decrease in wholesale price index implies deficient demand. It is a situation in which demand is less than supply.

(2) **Estimation of Monetary Value and Real Value:** The wholesale price index can be used to estimate the monetary value and real value of aggregates like national income and expenditure. Monetary value is the value estimated at current year prices. Real value is the value estimated at base year prices or at constant prices. The monetary aggregate can be **converted into real aggregate by applying the following formula:**
 Real Aggregate of the Current Year = Monetary Aggregate of the Current Year

Price Index of Base Year/Price Index of Current Year

- (3) **Indicator of Rate of Inflation:** The wholesale price index is also applied to calculate the rate of inflation in a country. It refers to the rate at which prices tend to increase over time.

Assertion Reason Answer:

1. (b) Both assertion and reason are not true. The reason is not the correct explanation of the assertion.
2. (b) Both assertion and reason are not true. The reason is not the correct explanation of the assertion.

