## Unit - I
## Introduction to Statistics

**What do you understand by statistics?**

**Introduction**

The word statistics was first introduced in 18th century. It is derived from the words 'Status' in the Latin, 'Statista' in the Italian, 'Statistique' in the French and 'Statistik' in the German. It is used for collection and arrangement of facts or data. Statistics are used in various fields, the use of statistics is unavoidable. It can be expressed as follows:

"Statistics without other sciences has no root and other sciences without statistics bear no fruit"

**Define statistics?**

**Definition of statistics**

Statistics may be defined as the science of collection, presentation, analysis and interpretation of numerical data.

**Explain the collection of data?**

**Collection of data**

There are two major categories to collect a data, they are

- Primary data
- Secondary data

**Primary data**

Data which is collected by actual observation or count is called primary data. Which are collected by individuals or through by agents.

**Secondary data**

Data which are derived from past recoded (i.e, primary data) is called secondary data.

**What are the Methods to collect primary data?**

**Method of collecting primary data**

There five method to collect primary data, they are

- Direct personal interviews
- Indirect oral interviews
- Information from correspondents
- Mailed questionnaire method
- Schedules sent through enumerations

**Direct personal interviews**

In this method the investigator personally meets them and asks question to gather the necessary information.

*Merits*

- Original data are collected
- True and reliable data can be had
- Response will be more encouraging, because of personal approach
- A high degree of accuracy can be aimed
- The investigator can extract correct information
- Misinterpretations, if any, on the part of informant(from whom information's are collected ) can be avoided
- Uniformity and homogeneity can be maintained

*Demerits*

- It is un suitable when area is large
- It is expensive and time consuming
- The chances of bias are more
- An untrained investigator will not bring good result
- One has to collect information according to the convenience of the informant

**Indirect oral interview**

In this method the investigator approaches the witnesses or third party to collect the information. The investigator interviews the people, who are in directly or indirectly connected with the problem under study. Eg we are asked to collect information relating to the gambling or drinking habits of people.

*Merits*

- It is simple and convenient
- It saves labour, money and time
- It can be used in the investigation of a large area
- The information is unbiased
- Adequate information can be had

*Demerits*

- Absence of direct contact is there; the information cannot be relied
- Interview with an improper man will spoil the results
- In order to get the real position, a sufficient number of persons are to be interviewed
- The careless attitude of the informant will affect the degree of accuracy
- Witnesses may colour the information according to their interests

**Information through agencies**

In this method, local agents or correspondents will be appointed. They collect the information and transmit it to the office or person. They do this according to their own way s and tastes. Eg this method is used in newspapers, periodicals, ect. ,

*Merits*

- Extensive information can be had
- It is the most cheap and economical method
- Speedy information is possible
- It is useful where information is needed regularly

*Demerits*

- The information may be biased
- Degree of accuracy cannot be maintained
- Uniformity cannot be maintained
- Data may not be original

**Mailed questionnaires**

A set of well-define question are prepared. It is called questionnaires. This questionnaire is send to all the informants by post with the covering letter. The informants are requested to fillup the set of questions in the given blank space and return it within time.

*Merits*

- It is relatively cheap
- It is fast if the informants respond duly
- It is preferable when the informants are spread over a wide area

*Demerits*

- The informants should be literates to answer the question
- There is possibility of non-respondents ie not sending back the filled questionnaire
- People may not give the correct answer and thus one is led to false conclusion
- The questionnaire is inelastic. Asking supplementary question is not possible
- Sometimes the informants may not be willing to give written answers, apart form causing delay

**Schedules send through enumerators**

In this method, no of enumerators are selected and trained. The trained person goes to the informants along with the questionnaire and gets replies to the questions in the schedule and records their answers. This method is widely used in the collection of primary data.

*Merits*

- This method is very useful in extensive enquiries
- It yields reliable and accurate results, because enumerators are educated and trained
- The scope of the enquiry can also be greatly enlarged

- Even if the respondents are illiterate, this technique can be widely used
- As the enumerations personally obtain the information, there is less chance of non-response.

*Demerits*

- This is a very costly method, as the enumerators are trained and paid for
- This method is time-consuming, because the enumerators go personally to obtain the information
- Personal bias of the enumerators may lead to false conclusion
- The quality of the collected data depends upon the personal qualities of the enumerator
- It is not suited to all persons due to its costliness

**What are the sources of secondary data?**

**Sources of secondary data**

There are various sources of secondary data. It can be divided in to two broad categories, they are

- Published sources
- Unpublished sources

**Published source**

There are several published sources, they are as follows

- *International publications*
- *Official publications of central and state governments*
- *Semi-official publications*
- *Publications of research institutions*
- *Publications of commercial and financial institutions*
- *Reports of various committees and commissions appointed by the government*
- *Journals and newspapers*

**Unpublished source**

There are various sources of unpublished data. They are the records maintained by various government and private offices, the researches carried out by individual research scholars in the universities or research institutes.

**What are the precautions to take when using secondary data?**

**Precautions in the use of secondary data**

Secondary data is one which has been collected by someone for some investigation and is used by someone else for another investigation. To collect secondary data we have to consider following precautions,

- Suitability
- Adequacy
- Reliability

**Suitability**

The data chosen should be suitable for the present study. It can be judged by the nature and scope of the present enquiry with the original enquiry. For example, if the objective of present enquiry is to study the fly fever and if the data provide only blood pressure, such a data are unsuitable.

**Adequacy**

If the data are suitable for the purpose of investigation, then we must consider whether the data are useful for the present analysis, that is, the data contain all the information of that period and from all the stipulated units

**Reliability**

The reliability of the data can be tested by finding out the agency that collected such data. If the agency has used proper methods in collecting data, statistics may rely upon ie we have to check whether properly trained enumerators are used, whether there was proper check at each of the stages of collection and classification.

**What do you mean by classification?**

**Classification of data**

Classification is the process of arranging the available facts into homogeneous groups or classes according to resemblances and similarities.

**Define classification?**

**Definition**

The process of grouping a large number of individual facts or observations on the basis of similarity among the items is called classification.

**Give the characteristic of the classification?**

**Chief characteristic of the classification**

- All the facts are classified into homogeneous groups by the process of classification.
- The basis of classification is unity in diversity.
- Classification may be either real or imaginary.
- The classification may be according to either similarities or dissimilarities.
- It should be flexible to accommodate adjustments.

**What are the objects of the classification?**

**Objects of classification**

The chief objectives of classification are

- To condense the mass of data.
- To presents the fact in simple form.
- To bring out clearly the points of similarity and dissimilarity.
- To facilitate the comparison.
- To bring out the relationship.
- To prepare data for tabulation.
- To facilitate the statistical treatment of the data
- To facilitate easy interpretation.
- To eliminate unnecessary details.

**What are the rules of classification?**

**Rules of classification**

The following are the rules for classification, it should posses the following guiding principles:

- Exactness
- Mutually exclusive
- Stability
- Flexibility
- Suitability
- Homogeneity
- Mathematical accuracy

**What are the types of classification?**

**Types of classification**

There are four types of classification, they are

- Geographical classification
- Chronological classification
- Qualitative classification
- Quantitative classification

*Geographical classification*

It is also called spatial classification. In this data are classified into area-wise or location-wise i.e, state, districts, cities, town, etc.,

For example

The awareness of TB among the town peoples are given below

| Name of the towns | No of persons |
|---|---|
| Chennai | 100000 |
| Tiruchi | 80000 |
| Coimbatore | 85000 |
| Salem | 56000 |

*Chronological classification*

In this type data are classified into time-wise or historical-wise. That is it is based on year, months, weeks, days, hours, etc.

For example

| Year | Population in million |
|---|---|
| 1921 | 248 |
| 1931 | 276 |
| 1941 | 313 |
| 1951 | 489 |

*Qualitative classification*

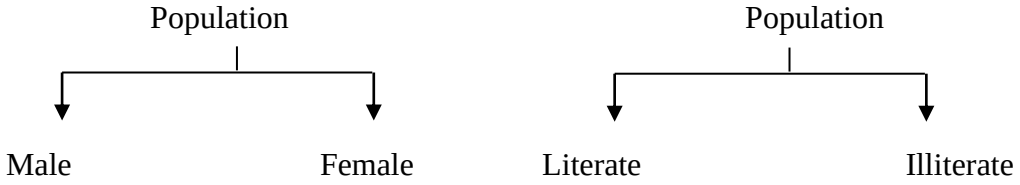
When the data are classified according to some quality or attributes, such as sex, honesty, intelligence, literacy, blindness, colour, deafness, religion, marital status, etc the classification is termed as qualitative or descriptive attributes. There are two type of the qualitative classification, they are

- Simple classification
- Manifold classification

*Simple classification*

If the data are classified into only two classes, such as literate and illiterate and honest and dishonest etc.,

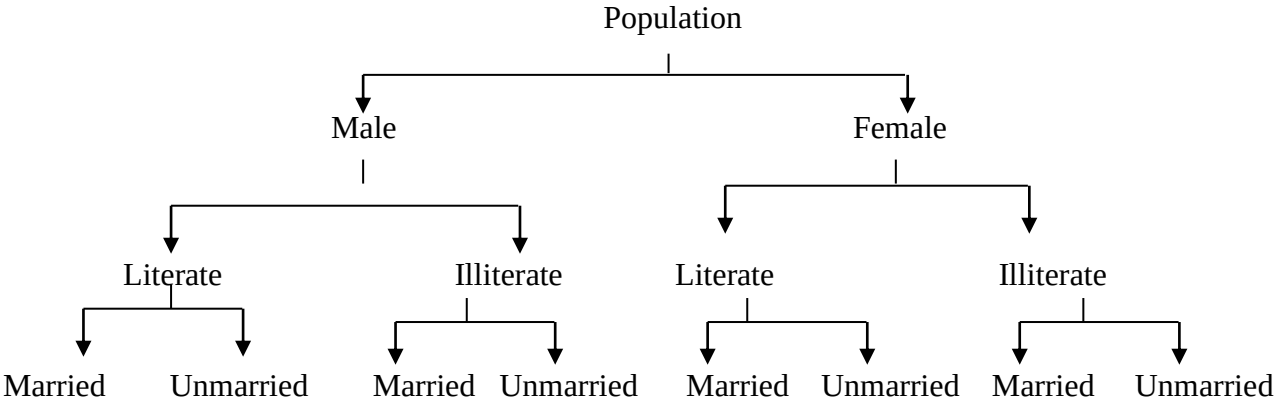
For example




*Manifold classification*

In this classification, data are classified into more than one attribute then the classification is called manifold classification.

For example




**Quantitative classification**

If the data are classified according to some characteristic which is capable of quantitative measurement like age, income, height, weight, etc.

For example

| Marks | No of students |
|---|---|
| 10-20 | 10 |
| 20-30 | 13 |
| 30-40 | 7 |
| 40-50 | 12 |

**Tabulation of data**

**Define tabulation?**

**Definition**

Tabulations is the process of arranging data systematically in rows and columns of a table.

**What are the objectives of tabulation?**

**Objective of Tabulations**

- Large and complex data can be presented in a neat and compact form.
- Nature of data can be easily understood.
- Much of time is saved by looking the data
- The table is easier to proper comparison.
- A table facilitates further analysis of data.
- The table is convenient form for diagrammatic representation of data.
- To present the fact in minimum space.
- To detect the errors and omission in the data.

**Explain the parts of tabulation?**

**Parts of tabulation**

A good statistical table is an art. The following parts must be present in all tables:

- Table number
- Title
- Head note
- Caption
- Stubs
- Body of the table
- Foot note
- Source note

**Table number**

A table should always be numbered for identification and reference in the future.

**Title**

Each table should be given a suitable title. It must be written on the top of table. It must describe the contents of the table. It must explain

- What the data are
- Where the data are
- Time or period of data
- How the data are classified, etc

**Head note**

It is the statement, given below the title and enclosed in brackets. For example the unit of the measurement is written as a head note, such as ‘in millions’.

**Captions**

There are the headings for the vertical columns. They must be brief and self-explanatory. They have main heading and sub heading and must be write in small letters.

**Stubs**

These are the headings for the horizontal rows. Stubs are wider than columns.

**Body of the table**

It contains the numerical information. It is the most important part of the table. The arrangement in the body is generally from left or right in rows and from top to bottom in columns.

**Foot note**

If any explanation regarding any item is necessary, foot notes should be given.

**Source note**

It refers to the source from where information has been taken. It is useful to the reader to check the figures and gather additional information.

**What are the types of table?**

**Types of tables**

Statistical tables can be classified into a number of ways. There are many categories depending upon:

- On the basis of coverage
- On the basis of objective(purpose)
- On the basis of originality

*On the basis of coverage*

It is further divided into

- Simple table
- Complex table

Simple table

In this table the data are classified according to only one characteristic. It is also called one – way table.

For eg:

Distribution of Marks

| Class marks | No of students |
|---|---|
| 20-30 | 10 |
| 30-40 | 18 |
| 40-50 | 22 |
| Total | 50 |

Complex table

In this table the data are classified into more than one characteristic. It is more popular, because it helps appropriate consideration of all related facts. It may takes two – way, three – way and manifold table.

For eg:

Distribution of marks (boys & girls)

| Class marks | Number of students | | Total |
|---|---|---|---|
| | Girls | Boys | |
| 20-30 | 10 | 15 | 25 |
| 30-40 | 15 | 10 | 25 |
| Total | 25 | 25 | 50 |

*On the basis of objective*

It is further divide into

- General purpose table
- Special purpose table

General purpose table

It is also called informative table and provides information for general use. It is ordered in chronologically.

For eg:

The Census reports.

Special purpose table

It is also called a summary table or text table or analytical table or derivative table or derived table. It presents the data relating to a particular purpose.

For eg:

Ratios, percentages, etc.

*On the basis of originality*

The statistical table may be classified into:

- Primary table
- Derived table

Primary table

It contains actual and absolute facts.

Derived table

It is derived from primary table.

For eg

Average, dispersion, correlation, etc.

**Give the difference between classification and tabulation?**

**Difference between Classification and Tabulation**

The following are different between them:

| Classification | Tabulation |
|---|---|
| 1. It is process of dividing the data into homogeneous subgroups.<br>2. It is the condensed form of mass data.<br>3. This the process of analysis of data.<br>4. Careful planning for tabulation is necessary even at this stage. | 1. It is the process of arranging the classified into rows and columns<br>2. It is the permanent form and used to refer the data readily.<br>3. This the process of presentation of data.<br>4. This is mechanical function after classification. |

**Frequency Distribution**

**What are the types of frequency distribution?**

Frequency distribution is basically divided into two types

- Uni-variate frequency distribution
- Bi-variate frequency distribution (Two-way)

*Uni-variate frequency distribution*

It is further divided into

- Individual observation
- Discrete frequency distribution
- Continuous frequency distribution

## MEASURES OF CENTRAL TENDENCY

**Define average?**

**Definition**

Average is a value which is typical or representative of a set of data.

**What are the objectives of average?**

**Objectives of average**

- Average provide a quick under standing of complex data
- Averages enable comparison
- Average facilitate sampling techniques
- Average helps to further statistical analysis
- Average establish the relationship between variables

**Give the Characteristics of average?**

**Characteristics of average**

- It should be rigidly defined
- It should be based on all the items
- It should not be unduly affected by extreme items
- It should be capable of further algebraic treatment
- It should be capable of being used in further statistical computation
- It should be simple to understand and easy to calculate
- It should have sampling stability

**What are the Types of averages?**

**Types of averages**

There are five types of averages, they are

- Arithmetic mean
- Median
- Mode

- Geometric mean
- Harmonic mean

**Mean**

**Define mean?**

Definition

Arithmetic mean is the total of the values of the items divided by their number.

**Give the merits and demerits of mean?**

Merits

- It is easy to understand.
- It is easy to calculate
- It is used in further calculation
- It is rigidly define
- It is based on the value of every item in the series
- It provides a good basis for comparison
- The mean is more stable measure of central tendency

Demerits

- The is unduly affected by the extreme items
- It is unrealistic
- It may lead to a false conclusion
- It can not be accurately determined even if one of the values is not known
- It is not useful for the study of qualities like intelligence, honesty and character
- It can not be located by observation or the graphic method
- It gives greater importance to bigger item of series and lesser importance to smaller items

**Median**

**Define median?**

Definition

Median is the value of the middle most item when all the items are in the order of magnitude.

**Give the merits and demerits of median?**

Merits

- It is easy to understand and easy to compute
- It is quite rigidly define
- It eliminates the effect of extreme items
- It is open to further algebraic process
- Science it is positional average, median can be computed even if the items at the extremes are unknown

Demerits

- It is more effected by fluctuations of sampling than in mean
- In case of continuous series, the median is estimated, but not calculated
- When the item is large, arrange the item is difficult process
- It ignores the extreme items.
- If the distribution of item is irregular, the typical representative of the observations cannot be computed.

**Mode**

**Define Mode?**

Definition

Mode is the value which has the greatest frequency density.

**Give the merits and demerits of mode?**

Merits

- Mode is not unduly affected by extreme values.
- It is simple to understand and easy to calculate.
- It can be calculated even from open-end data.
- It can be determined graphically.

- It is usually an actual value as it occurs most frequently in the series.
- It is most representative average.

Demerits

- It is not suitable for further mathematical treatment.
- It may not give weight to extreme items.
- It will not give the aggregate value as in average.
- It is stable only when the sample is large.
- Mode is influenced by magnitude of the class-interval.

**Geometric mean**

**Define Geometric mean?**

Geometric mean of N values is the $N^{th}$ root of the product of the N value.

**Give the merits and demerits of geometric mean?**

Merits

- Geometric mean is highly useful in averaging ratios, percentages and rate of increase between two periods.
- Geometric mean is important in the construction of index numbers.
- It based on magnitudes of all the items.
- It is used in business, economic and social data.

Demerits

- It has less sampling stability than the A.M.
- It cannot be calculated for open – end data
- It cannot be found graphically.
- It is not only defined for qualities
- It cannot be calculated even if one value or one mid value is zero or negative.

**Harmonic mean**

**Define Harmonic mean?**

Definition

Harmonic mean is the reciprocal of the mean of the reciprocals of the values.

**Give the merits and demerits of Harmonic mean?**

Merits

- It is rigidly defined.
- It is based on all the observations of the series.
- It is suitable in case of series having wide dispersion.
- It is suitable for further mathematical treatment.
- It gives less weight to large items and more weight to small items.

Demerits

- It is difficult to understand.
- All the values must be available for computation.
- It is not popular.
- It is usually a value which does not exist in series.

**End of the unit – I**

---

**UNIT-II**

## MEASURES OF DISPERSION

**Define Measure of Dispersion?**
**Definition**

"Dispersion is the measure of the variation of the items".

**What are the methods of measuring the Dispersion?**
**Methods for Measuring Dispersion**

The following are the important methods of studying variation.

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

**Range**
**Define Range?**
Definition

Range is the difference between the greatest(largest) and the smallest of the values.

Range = L – S
L - Largest value
S - Smallest value.

**Give the merits and demerits of Range?**
Merits

- It is simple to understand.
- It gives a quick answer.

Demerits

- It cannot be applied to open end class intervals.
- It is not suitable for mathematical treatment.

**Quartile deviation**
**Define quartile deviation?**
Definition

Quartile Deviation is half of the difference between the first and the third quartiles.

**Give the merits and demerits of quartile deviation?**
Merits

- It is simple to understand and easy to calculate.
- It can be calculated for data with open end classes also.

Demerits

- It ignores the first 25% of the items and the last 25% of the items.
- It is a positional average, hence not amenable to further mathematical treatment.
- It gives only a rough measure.
- Like range, it does not measure the deviation about any measure of central tendency.

**Mean Deviation**
**Define mean deviation?**
Definition

Mean deviation is the arithmetic mean of the absolute deviations of the values about their arithmetic mean or median or mode.

**Give the merits and demerits of mean deviation?**

Merits

- They are based on all the items.
- They are simple to understand and not difficult to calculate.
- They do not vary much from sample to sample.
- Mean deviations are rigidly defined.

Demerits

- It is not widely used in business or economics.
- It is a non-algebraic treatment.
- They could not be manipulated. Combined means deviation could not be found.

**Standard Deviation**

**Define Standard deviation?**

Definition

Standard Deviation is the root mean square deviation of the values from their arithmetic mean.

**Give the merits and demerits of standard deviation?**

Merits

- It is rigidly defined.
- S.D. is used in finding coefficient of variance.
- It can be used to calculate the combined standard deviation.
- The standard deviation provides the unit of measurement for the normal distribution.

Demerits

- It is not easy to understand, and it is difficult to calculate.
- It gives more weight to extreme values.
- It cannot be used for the purpose of comparison..
- It is affected by the value of every item in the series.

**Skewness**

**Define Skewness?**

Definition

Skewness is the degree of asymmetry, or departure from symmetry of a distribution.

**End of the unit – II**

---

**UNIT-III**

# CORRELATION

**Define correlation?**

**Definition**

Correlation analysis attempts to determine the degree of relationship between variables.

**Explain the types of correlation?**

**Types of Correlation**

Correlation is classified into many types, but the important are:

- Positive and negative correlation
- Simple and multiple correlation
- Partial and total correlation
- Linear and non-linear correlation.

**Positive and Negative correlation:**

**Positive correlation:**

If two variables tend to move together in the same direction, then the correlation is called positive correlation.

Example

Height and weight

Rainfall and yield of crops.

**Negative correlation**:

If two variables tend to move together in opposite directions, then the correlation is called negative correlation.

Example

(i) Price and demand.

(ii) Yield of crops and price.

**Simple and multiple correlations**

**Simple correlation**

In simple correlation we study the relationship between two variables only.

Example

(i) Demand and price

(ii) Sales and production.

**Multiple correlations**

In multiple correlations we study the relationship between more than two variables.

Example:

(i) Price, demand and supply.

(ii) Production, sales and income.

**Partial and Total correlation**

**Partial correlation**

The study of two variables excluding some other variables is called partial correlation.

Example

We study price and demand, eliminating the supply side.

**Total correlation**

In total correlation, we study all the variables are into account.

**Linear and non-linear correlation**

**Linear correlation**

If the ratio of change between two variables is uniform, then there will be linear correlation between them.

**Non-linear correlation**

If the ratio of change between two variables is not uniform, then there will be non-linear correlation.

**Give the uses of correlation?**

**Uses of Correlation**

1. Correlation is useful in physical and social sciences.
2. Correlation is very useful to economists to study the relationship between variables, like price and demand etc.,
3. Correlation is useful to businessman to estimate costs, sales, price and other related variables.

4. Sampling error can also be estimated
5. The relation between variables can be verified and tested for significance with the help of the correlation analysis. The effect of correlation is to reduce the range of uncertainty of our prediction.
6. The co-efficient of correlation is a relative measure and we can compare the relationship between variables which are expressed in different units.
7. Correlation is the basis for the concept of regression and ratio of variation.

**What are the methods of finding the correlation?**

**Methods of Correlation**

The different methods of finding out the relationship between two variables are

Graphical method

Scatter Diagram or Scatter gram

Simple Graph or Correlogram

Mathematical method

- Karl – Pearson's Coefficient of Correlation
- Spearman's Rank Coefficient of Correlation
- Coefficient of Concurrent Deviation

**Give the merits and demerits of scatter diagram?**

Merits

Scatter diagram is a simple and attractive method of finding out the nature of correlation between two variables.

It is the non mathematical method of studying the correlation. It is easy to understand.

We can get rough idea at a glance whether it is the positive or negative correlation.

It is not influenced by extreme items.

It is a first step in finding out the relationship between two variables.

Demerits

By this method we cannot get the exact degree or correlation between two variables. It gives only a rough idea.

**Give the merits and demerits of Karl – Pearson's coefficient of correlation?**

Merits

- Karl – Pearson's correlation coefficient is the most popular correlation coefficient. It is used in regression equation also.
- It is superior to other methods.
- It can be calculated directly using the each and every pairs of values.
- The population correlation coefficient can be estimated from the sample values.
- The signification of the sample correlation coefficient can be tested.

Demerits

- The correlation coefficient is unduly affected by extreme values.
- From the value of r, it cannot be known whether the assumption of linear relationship between the variables holds or not.
- Karl –Pearson's coefficient of correlation is difficult to compare.
- It is often misinterpreted.

**Give the merits and demerits of Spearman's rank correlation coefficient?**

Merits

- It is useful in qualitative analysis.
- It is the only method when ranks are given
- It can also be calculated when the values of the variable are given.
- It is simple to understand and easy to calculate

Demerits

- It is difficult to rank the item when 'N' is large.
- It cannot be used to calculate bi – variate frequency table.

**Give the merits and demerits of the concurrent deviation?**

Merits

- It is easy to calculate even if 'N' is large and the values are large.
- It is simple to understand.
- It is easy to calculate.

Demerits

- It just give the rough idea about the existing correlation between two variables.
- It does not consider the quantum of deviations.
- It is not used in bi- variate frequency table.
- It considers neither the values nor their relative positions.

## REGRESSION

**Define Regression?**

**Definition**

Regression is the measure of the average relationship between two or more variable in terms of the original units of the data.

**Give the use of regression?**

**Uses of Regression**

1. Regression analysis is used in Statistics and other disciplines.
2. The regression line equation helps to estimate the value of dependent variable from the values of independent variables.
3. It is widely used in social sciences like economics, natural and physical sciences.
4. We can calculate coefficient correlation ( r ) and coefficient of determination ( $r^2$ ) with help of regression coefficient.
5. Regression analysis predicts the value of dependent variables from the values of independent variable.

**Give the difference between correlations and regression coefficient?**
**Difference between correlations and regression**

| CORRELATION | REGRESSION |
|---|---|
| 1.Correlation is the relationship between two or more variable, which vary in sympathy with the other in the same or the opposite direction. | 1.Regression means going back and it is a mathematical measure showing the average relationship between two variables. |
| 2.Both variables x and y are random variables. | 2. Here x is a random variable and y is a fixed variable. Sometimes both the variables may be random variables. |
| 3.There may be nonsense correlation between two variables. | 3. In regression there is no such nonsense regression. |
| 4.It has limited application, because it is confined only to linear relationship between the variables. | 4.It has wider application, as it studies linear and non-linear relationship between two variables. |
| 5.The coefficient correlation is a relative measure. The range of relationship lies between ±1. | 5. Regression coefficient is an absolute figure. If we know the value of the independent variable , We can find the value of the dependent variable. |
| 6. It is not very useful for further mathematical model treatment. | 6. It is widely used for further for further mathematical treatment. |
| 7. If the coefficient of correlation is positive, then the two variables are positively correlated and vice versa. | 7. The regression coefficient explains that the decrease in one variable is associated with the increase in the other variable. |

**End of the unit – III**

## Unit - IV
## Index Numbers

**Define index numbers?**

**Definition**

An index number is a statistical measure designed to show changes in a variables or group of related variables with respect to time, geographical location or other characteristics.

**Give the Characteristics of index numbers?**

**Characteristics of index numbers**

According to the definition of index numbers the following characteristic are important:

- **Index numbers are specialized averages**

  Commonly averages are used to compare the variables which are expressed in the same units. But index numbers are used to compare the variables which are expressed in different units. Therefore it is called specialized average.
- **Index numbers are expressed in the percentages**

  To show the extent of changes it is expressed in percentages.
- **Index numbers are not capable of measuring the direct measurement.**
- **Index numbers are for comparison**

  The index numbers by their nature are comparative. They compare changes taking place over time or between places and likely categories.

**Give the uses of index numbers?**

**Uses**

Index numbers are most widely used statistical device.

- It measure the relative changes. They give the better idea of changes in level of prices, production, business activity, employment etc.
- Index numbers are useful to better comparison.
- Index numbers are good guides.
- Index numbers are Economic barometers.
- Index numbers of general price level will measure the purchasing power of money.
- Index numbers are used in wage adjuster. It is done with the help of consumers' price index numbers. For example it is use to determine the dearness allowances of the employees.
- It is useful to compare the standard of living.
- It is helpful in formulating policies.

**Give the limitations of index numbers?**

**Limitations of index numbers**

Even though index numbers are very important in business and economic activities, they have their own limitations, they are

- They may be error in each stage of the construction of the index number, namely selection of commodities, selection of base year, selection of weight, etc.
- Index number may not represent the exact change in price level, because they are based on sample data.
- Tastes, habit and customs of people changes in course of time and may make the weighting not suitable for the present data.
- In each index there is an index error, because there is no formula for measuring yhe price changes. So there is formula error.
- By selecting a suitable year as the base year, selfish persons may get their desired results.

**What are the types of index numbers?**

**Types of index numbers**

There are various types of index numbers, but in brief we see three kinds and they are

- Price index numbers
- Quantity index numbers
- Value index numbers

*Price index numbers*

It is a index number which compares the prices for a group of commodities at a certain time or at a place with prices of a base year or period. There are wholesale price index numbers and retail price index numbers. The wholesale price index reveals the changes in the general price level of a country. Retail price index reveals the changes in the retail price of commodities, such as, consumption goods, bank deposits, bonds, etc.

*Quantity index numbers*

Quantity index numbers study the changes in the volume of goods produced or consumed. For example industrial production, agricultural productions, import, export, etc.

*Value index numbers*

It compares the total value of a certain period with the total value of the base period. For example, it gives the indices of profits, sales, inventories, etc.

**Explain the Problems in the construction of index numbers?**

**Problem in construction of index numbers**

The following are important points to construct the index numbers

- Purpose or object

  In this one should clearly determine the purpose for which the index numbers are to be constructed, because there is no all purpose index number.

  For eg:

  If we want to study the changes in the value of money, then we have to construct index numbers of wholesale price.

- Selection of base

  It is very important point in construction of index numbers. Every index numbers must have a base. One cannot say whether the price level has increased or decreased, unless one compare the price level of current year with the price level of previous year.

  The selection of base period or year must be recent and normal. A normal year is free from economic and natural disturbances, widespread failure of rains, earthquakes, war, strikes, production crisis, etc.

  The base has the following types

  - Fix base
  - Average base
  - Chain base

- Selection of commodities

  If we study the price changes of one commodity, we have to include only one item.

  For eg:

  If we study the changes in production of cloth, then we may include the production of mill cloth, power loom cloth, handloom cloth, silk, khadi, etc.

  Consider other eg:

  We are interest to study the cost of living index number of low income group, we have to select only those item or commodities, which are mostly consumed by that group

- Sources of data

Here price of commodities or items are the source of data. The prices of commodities are the raw materials for the construction of index numbers. The prices may be collected from the public sources or from the standard commercial magazines.
For eg:
We study the changes in industrial production, we must collect the prices relating to the production of various goods of factories.

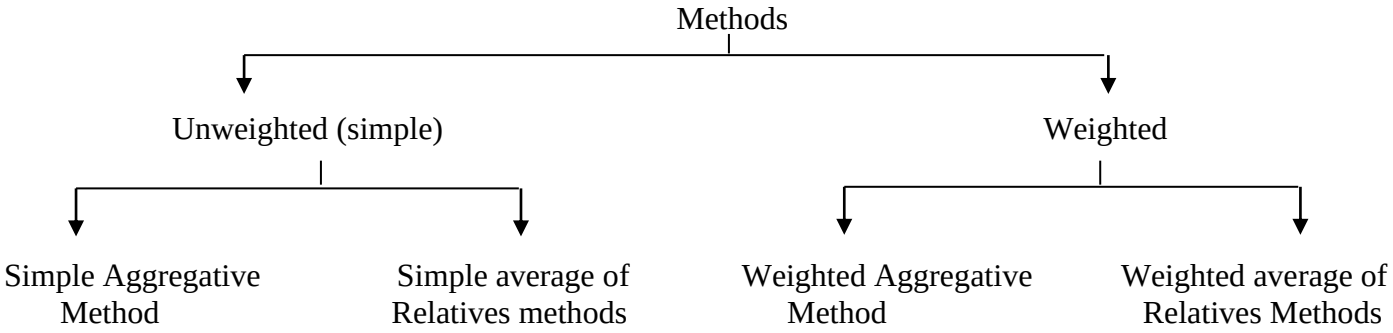
- Selection of averages

  One can use any averages. But in practice, the arithmetic average is used, because it is easy for computation; geometric mean and harmonic mean are difficult to calculate. But geometric mean is preferred because of the following characteristics. i) Geometric mean is the best measure and ii) It gives less weight to bigger items and more weight to smaller items.

- Weighing

  All commodities are not equally important. The main purpose of an index number of prices is to ascertain the changes in the price level. In case of simple average, all the commodities will have equal importance. In practice it is not possible.
  For eg:
  When the price of rice or wheat is doubled and the price of sugar is halved, the people suffered much, because the price of rice which is essential has been doubled.
  There are two methods of weighting
  - Implicit weighting
  - Explicit weighting

**What are methods to calculate the Index numbers?**
**Methods of Index Numbers**




## Cost of living index number

**What is meant by Cost of living index number?**
**Cost of living index number**

Cost of living index number is also called consumer price index number. It can be used in the appropriate place. It is design to measure the average change over time in the price paid by ultimate consumer for a specified quantity of goods and services. It helps to measure the change in the cost of living of the workers due to change in the retail price. A change in the price level affects the cost of living of different classes of people differently. The general index numbers fails to reveal this. So there is need to construct consumer price index. People consume different types of commodities. People's consumption is also different from man to man, place to place, class to class i.e., richer class, middle class and poor class.

**What are the needs or scope of Cost of living index number?**

**Scope of Cost of living index number**

The scope of the consumer price index is necessary to specify the population groups covered; for example: working class. Middle class, poor class, richer class, etc. and the geographical areas must be covered as urban, rural, city, town etc.

**What are uses of consumer prices index number?**

**Uses of consumer prices index number**

- This is very useful in the wage negotiations and wage contracts and allowance adjustment in many countries.
- Government can make use of these indices for wage policy, price policy, taxation, general economic policies and rent control.
- Changes in the purchasing power of money and real income can be measured.
- We can analysis the market price for particular kinds of goods and services by this index.

**Explain the construction of consumer price index number?**

**Construction of consumer price index number**

There are some precautions in to be taken in the construction of price index.

**Determination of class of people**

It is very important to know for whom the index is going to frame, i.e., rich or poor, rural or urban. It it should be carefully determined. We must clearly define scope of the index.

For example

When we consider the consumption habit of the working class,we must refer to industrial workers or agricultural workers.

**Selection of base period**

The base period must be the period of economic stability. Fluctuations due to changes in the season may be elimated by taking the completed year as the base.

**Conducting family budget enquiry**

It cover the population groups for whom the index is meant. The object of it is to determine amount spend on different items of consumption by the average group in a population group. We must take quantity of commodities consumed and the price level of each commodity. Thus we can find out the consumption pattern. Family budget survey is based upon a random sample basis, where as the families are selected by lots.

The expenditure on different items is classified in to broad heads and in certain well – defined groups. They are:

- Food
- Clothing
- Fuel and Lighting
- House rent
- Miscellaneous

**Collection of price quotations**

Collection of price quotation is very important in the construction of costing of living index price which vary from place to place, to shop, and from consumer to consumer and it is difficult and time consuming job to collect the retail price. The principal to be adopted in the collection retail prices are:

- Price must be related to a fixed list of item for a fixed quality.
- Retail price must be the price which is given by the consumer.
- If discount is given to all customers it can be taken in to account.
- We must take in to account the ration price or control price and open market price.

In the cost of living index, price is the most important element. It is collected by special agent or mail questionnaires. To do this first we must select the special agents and give the special training to

them. They must be instructed properly and give the list of items to be priced. Price verification must be conducted by 'check pricing' or 'purchase checking'.

**What are the methods to construct consumer price index?**
**Methods of constructing consumer price index**

There are two method of construction, they are:

- Aggregate expenditure method or aggregate method.
- Family budge method or method of weighed relative.

**What are the precautions in the use of cost of living index numbers?**
**Precautions in the use of cost of living index numbers**

Many times, it is misinterpreted. Therefore, are the points to be kept in the mind while using these indices.

- Cost of living index number only measure changes in retail price from one period to another period. It does not tell us anything about variations in the living standard at different places.
- Weights can not be representative and if it is so index numbers would give misleading conclusions.
- The index numbers do not take in to account the changes in qualities.

**Give the difference between chain base and fixed base index numbers?**
**Difference between chain base and fixed base index numbers**

| Chain base | Fixed base |
|---|---|
| 1. The base year changes<br>2. Link relative method is used<br>3. Introduction and deletion of items are easy to calculate, without recalculation of entire series.<br>4. The calculation are tedious.<br>5. It is difficult to understand<br>6. It cannot be calculated if any one year are missing<br>7. It is suitable for short period only<br>8. Weight can be adjusted as frequently as possible.<br>9. Index number is wrong if the error in calculation of link relative. | 1. The base does not change<br>2. Link relative method is not used<br>3. Any changes in the commodities, will involve the entire index number to be recast.<br>4. The calculation is simple.<br>5. It is simple to understand<br>6. It can be calculated if any one year are missing.<br>7. It is suitable for long period only<br>8. Weight cannot be adjusted so frequently<br>9. In this, error is confined to the index of that year only. |

**End of the unit – I V**

## Unit –V
## ANALYSIS OF TIME SERIES

**Define Time Series?**

**Definition**

A time series is a set of observations arranged in chronological order.

**Give the uses of Time Series?**

**Uses**

The series is used in following field.

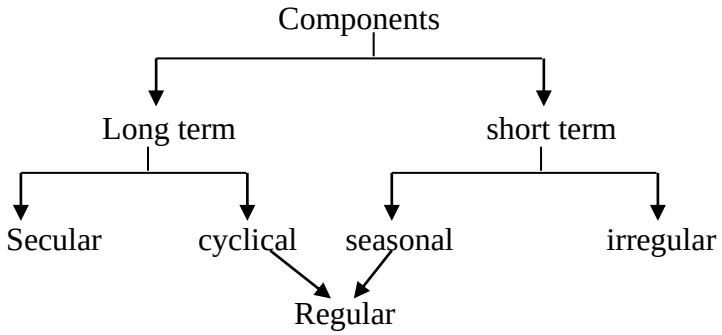
- Economics
- Business
- State administration and planning
- Science
- Astronomy
- Sociology
- Biology
- Research work etc., because of the following reasons

- It helps in understanding past behaviors and it will help in estimating the future behavior
- It helps in planning and forecasting.
- Comparison between data of one period with that of another period is possible.

**Explain the components of Time Series?**

**Components of Time Series**

There are four basic types of variations and these are called the components of time series




**Secular Trend**

Secular trend is also called long term trend. The general tendency of a series is to increase or decrease over a period of time.

**Increasing tendency is observed in,**

- Population
- Price
- Literacy etc.,

**Decreasing tendency is observed in,**

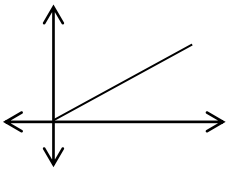
- Birth rate
- Death rate
- Poverty
- Illiteracy etc.,

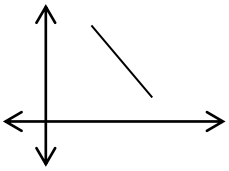
It is very rare to find a time series which neither increases nor decreases.

**Types of Trend**

**(i)Linear trend or straight line trend**

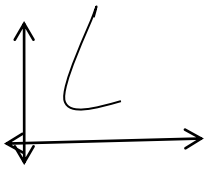
**Increasing trend Line** **Decreasing trend line**

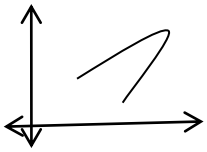




**(ii) Non-linear or curvilinear trend**

**Increasing trend curve** **Decreasing trend curve**





**Seasonal variation**

Seasonal variation is also called short-term variation (fluctuation).in this case variation occurs in short-term i.e., weekly, monthly, (or) quarterly.

The seasonal variation may occur due to,

a) **climate and natural forces**

In rainy season, sales of umbrellas are increased. In winter season, sale of woolen cloths will increase. In hot season sale of ice-cream, fruit- salad etc., will increase. Thus climate and weather play an important role in seasonal variation.

b) **customs and habits**

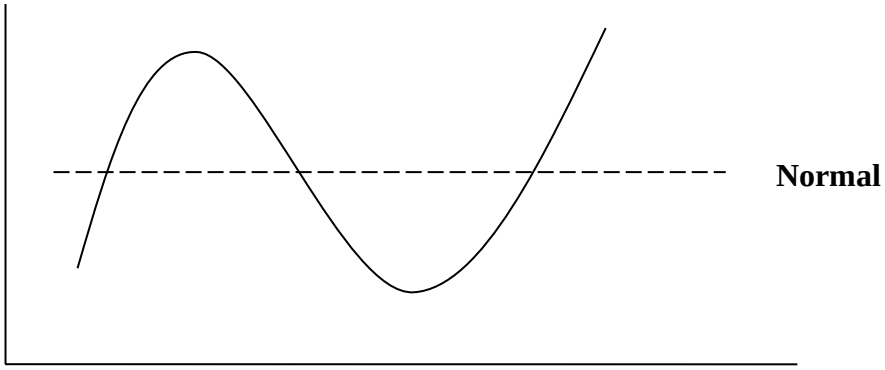
There is the customs of wearing new clothes, preparing sweets, and buying crackers for deepavali, onam, ramzan, chritmas, etc., at that time, there is more demand for clothes, sweets, and crackers. It will happen every year. in marriage season; the price of gold will increase.

**Cyclical variation**

Up and down movements are different from seasonal fluctuations, in that they extend over longer period of time, usually two (or) more years. This variation is mostly occurring in economic and business field. There are four prosperity (boom), recession, depression and recovery. This movement is known as cyclical variation.

**Business cycle**

`




**Irregular variation**

This variation is not predictable and forecast. This variation occurs irregularly. For example fire, floods, earthquakes, wars, lock-outs, strikes, etc., It is also called random variation or erratic fluctuation.

**What are the methods to calculate long term variation or secular trend?**

**Secular Trend**

There are four methods to estimate the secular trend. They are

- Graphic method
- Method of semi – averages
- Method of Moving averages
- Method of Least Squares

**Explain about graphic method? Or Give the procedure to find Graphical method?**

**Graphical method**

It is also known as free – hand method. X – axis represent time and Y – axis represent the observed data. Corresponding to each pair of time and observed value, a point is marked on a graph sheet. After marking all such possible point, the best line is drawn. It is the trend line. All the marked points do not lie on line. Hence the line is drawn, such that the following three conditions are satisfied:

- The number of points above the line is equal to the numbers of points below the line, as far as possible.
- The sum of the vertical distances of the points above the line equals that of the points below the line.
- The sum of the squares of the vertical distances of all the points from the line is minimum.

It is not easy to draw such a line. But method of least squares provides such a line mathematically.

**Give the Merits and Demerits of Graphic method?**

**Merits**

- It is simple method
- It is flexible. Based on the positions of the points, trend line(linear) or trend curve (non – linear) can be drawn.
- It gives better estimates when it is used by an experienced statistician.

**Demerits**

- It is subjective. Difference persons get different trend lines (or trend curves).
- It is not relied for prediction because of its subjective character.
- It is not easy to draw a central line as the points do not lie on a line. It requires experience and careful approach.

**Explain about method of semi – averages? Or Give the procedure to find method of semi – averages?**

**Method of Semi – Averages**

The time series is considered. When there are even numbers of year, the middle most year and the arithmetic mean of the observed values are found out for each half.

When there are odd numbers of year, the middle most year and the corresponding observed value are omitted.

The middle most years and the arithmetic mean of the observed values are then found out for each half. Based on them two points are marked on a graph sheet. The two points are joined by a straight line which is extended on either side. It is the trend line. The trend at any point of time can be found from that line. Only two points are marked on a line. There is no difficulty in drawing the line along the two points.

**Give the Merits and Demerits of method of Semi – Average?**

**Merits**

- It is not a subjective method. Under the graphic method, different person may get different trend lines. Semi – average method gives the same two points to all the persons for drawing the trend line.
- It involves very simple calculations. It is easy to adopt.
- The trend at any point of time can be found.

**Demerits**

- It is not flexible. When the trend line is non – linear, we could not get such a trend curve by this method.
- It is based on arithmetic mean. The demerits of the arithmetic means are the demerit of this method.

- When the trend of even one half is a fraction such as 331/3, it is difficult to mark the point on a graph sheet. Like any other graphic method, this is not relied on. In that situation, the algebraic methods, method of moving average and method of least square are preferred.

**Explain about the method of moving average?**

This method of moving averages is one of the most useful methods of estimating trend. It is an algebraic method. Graph sheet is not used for calculating trend.

For a series there is only one arithmetic mean; there are many moving averages. Moving totals are found and they are divided by appropriate number to get the moving average. The following two cases arise:

- Period of moving average is an odd number
- Period of moving average is an even number

**Write down the procedure for period of moving average is an odd number?**

Moving totals are found and written against the middle most years. Each moving total is divided by the period of moving average and the corresponding moving average is found.

Moving average is the trend. If short – term fluctuation is required, trend is subtracted from the observed value.

Let a, b, c,… be the observed values. When 3 yearly moving averages are required, a + b + c, b + c + d, c + d + e,… are the moving totals corresponding to second , third, fourth,… years. Each total is then divided by 3 to get the moving average.

That is $\frac{a+b+c}{3}$ $\frac{b+c+d}{3}$ $\frac{c+d+e}{3}$

are moving averages corresponding to second, third, fourth,…. Years. There is no moving total and moving average to the first and last year.

Similarly we can find for 5 yearly,7 yearly, 9yearly,…moving averages are calculate in a similar manner

**Write down the procedure for period of moving average is an even number?**

The mid year of the moving totals are not the given year in this case. Hence,2 period moving totals are found. The given years are found to be the mid years of these totals. 2 period moving totals are divided by twice the period of the moving averages to get the centred moving averages. The centred moving are the trend values.

**Give the Merits and Demerits of method of moving averages?**

**Merits**

- It is the simple method. The calculations are easy.
- It is the objective method. For a problem every one get the same moving averages.
- When the data are available for a few more year; the earlier calculations need not be redone. A few more moving averages would be available.
- If the period of moving average coincides with that of cyclical fluctuation, the cyclical fluctuation is completely eliminated. Better estimates of trend are obtained.
- In one of the methods, seasonal variations are found on the basis of moving averages. It is most widely used method of estimating seasonal variations.
- It is highly suitable when there is considerable fluctuation in the data.

**Demerits**

- The main purpose of analysis of a time series is defected. It can not be forecast the trend of any future period. It cannot estimate trend of a few observed years also. 2n yearly (n = 1, 2, 3, …) centred moving average method does not give the centred moving averages of the first n years and the last n years. (2n + 1) yearly moving average method does not provide the moving averages of the first n years and the last n years.
- The calculation are tedious when the period of moving average is large and an even number.
- The period of moving average should suit the nature of the series. Or else, a distorted picture of the time series will emerge.

**Explain the method of Least of Squares?**

**Method of Least of Squares**

By taking the time (X) as independent variable and the observed values (Y) as the dependent variable, the trend line of the form Y = a + bX can be formed as discussed in the ' method of least squares'.

**Give the Merits and Demerits of method of Least Squares?**

**Merits**

- Method of least squares is an objective method. Every one has to get the same trend equation for a data.
- It is flexible. Linear and non – linear trend equations can be found.
- Unlike the method of moving averages, the trend of each given year is available by this method. Similarly, the trend at any other point of time, past or future, can be calculated.
- It is the best method. It is mathematical and impersonal. It is better than graphic method and the method of semi – average as it provides the equation of non – linear trend. It is better than the method of moving averages when the consideration is estimation of the trend of a future period.
- The trend line obtained by this method is called the line of best fit. $\sum (Y - Y_t) = 0$ and $\sum (Y - Y_t)^2$ is the least for the line.

**Demerits**

- It is neither simple nor easy. It requires more time than the other methods.
- If the data for one or more years is added to the earlier data, the calculations are to be redone from the beginning unlike in the method of moving averages.
- Extreme values affect the results unduly unlike in the method of moving averages.

**What are the methods to calculate seasonal variation or short term fluctuation?**

**Methods of Seasonal Variation**

There are four methods are used to estimate the seasonal variation. They are

- Method of Simple Average
- Method of Moving averages
  - Difference from the moving averages.
  - Ratio – to moving average.
- Ratio to Trend Method
- Method of Link relatives.

**Write down the procedure for calculating the Method of Simple Averages?**

**Method of Simple Average**

This method assumes absence of trend in a time series. The following are the steps for calculating simple average.

- The data are arranged season-wise in chronological order.
- For each season, the total of the seasonal values is found and called seasonal total.
- Each seasonal total is divided by number of years and seasonal average is obtained.
- The total and the average of the seasonal average are found. The average is called grand average.
- Seasonal index of every season is calculated as follows:

$$\text{Seasonal index} = \frac{\text{Seasonal average}}{\text{grand average}} \times 100$$

**What are the merits and demerits Method of Simple Average?**

**Merits**

- It is easiest method..
- It is simplest and least time consuming method.

**Demerits**

- It assumes the absence of trend in a time series. This assumption is not always true.
- It assumes that the averaging process eliminates the seasonal fluctuations. It is also not true.

**End of the unit – V**