SYNOPSIS FOR

MARWADI SIKSHA SAMITHI
R.G. KEDIA COLLEGE
(Affiliated to Osmania University, Recognized by AICTE)
3-1-336, Esamia Bazar, Opp. New Chaderghat Bridge, Hyderabad- 500027.

BUSINESS STATISTICS

For B.Com III Semester

For the july to December 2023 Batch

Prepared by:
M.A AHMED

UNIT – I

Introduction of statistics, origin of statistics, History of statistics
Definitions of statistic, statistical methods
Singular sense

Characteristic of method:
1] Observation and collection

2] Organization of data
3] Presentation
4] Interpretation

Statistics as plural sense and its characteristics:
a] It is an agreeable of facts b] It is
numerically expressed
c] A reasonable degree of accuracy must be kept while collecting data

Scopes of statistics are: a]
Statistical method
b] Applied statistics

Statistical method:
A] Observation and collection
b] Organization c]
Presentation d] Analysis
e] Presentation

Applied statistics: A]
Description statistics
b] Scientific statistics
c] Analytical statistics
d] Inferential statistics
e] Inductive statistics

Nature of statistics:
Statistics as a science and as an art Law
of statistics:
1] The law of statistical regularity
2] The law of inertial of large number

Functions of statistics:
1] Systematic collection and presentation of facts 2] implacable of mass figures
3] Facilitates comparison
4] Help in formation and testing hypo sis

5] Helps in prediction
6] Helps in policy formation and decision making
7] Enlarge individual's experience
8] Study relationship between remains phenomena
9] Measures uncertainty

Importance of statistics:
Imp of statistics to a] common man
b] Stale
c] In business management
d] 1) insurance ii) financial statistics iii) misresearch iv) quality assurance v) a/c and audit
Objectives of statistics
Limitation of statistics
Source of statistics

Primary data:
1] Primary data is collected by person conducting the statistical enquiry
2] It is raw data
3] It is most relevant to study being conducted 4] It involves huge costs.

Secondary data:
1] The secondary data is collected by person agencies other than the one conducting the enquiry
2] It is processed and hence, not original
3] It may or may not be directly relevant to the study
4] It is relatively less expensive and takes less time and effort

Advantages of primary data:
1] Primary data is very reliable as it is being collected first hand
2] Primary data is most suitable for the study as it is being collected for the purpose of the study
3] Since data is being directly collected, the scope for miss-interpretation or loss of data is minimal.

Limitations of primary data:

1] Primary data is vulnerable to manipulation. The person collecting the data can drop some of the response from the total responses collected to tilt the results in a particular direction
2] It is a very expensive proposition
3] It involves a lot of time and effort on part of the investigator.

Advantages of secondary data:
1] Secondary data is processed data. Various statistical techniques would have already been applied on raw data and hence, it is possible to arrive at conclusions based on secondary data
2] The reliability of secondary data is high if the source of such data is a reputed agency or organization
3] Secondary data saves huge efforts and time that is required for collecting primary data.

Limitations of secondary data:
1] Secondary data may not be directly relevant to the study being conducted
2] Secondary data is reliable as the reputation of the agency providing it 3] It is processed data and hence, would have already lost some of the forced of the original data collected.

UNIT – II

DIAGRAMMATIC AND GRAPHIC PRESENTATION

Importance of diagrams:
1] Diagrams reduce huge amount of data into simple figures. Thus it is the best option for making short duration presentations to top management, government, customer, etc that cannot go into details.
2] Diagrams bring out the essence of the underlying data with great clarity. A person going through the data my get lost in detail, but will get the message immediately if the same data is presented as a diagram.

Limitations of diagram:

1] Diagrams are used to convey a message. The information that can be meaningfully presented by way of a diagram is limited. If too many details are packed in one diagram, have will be loss of clarity.
2] Diagrams have limited capability to highlight small differences in large measurements. For example, while plotting annual household incomes that can range from 10,000 to 10 cores, the difference between persons with annual household income of 50,000 and 1, 00,000 will not get highlighted.

General rules for constructing diagrams:
1] Title: each diagram must be given a suitable title to convey the main idea it is intended to portray.
2scale: scale should be selected consistent with the size of observations to be displayed.
3] Proportion between width and height: a diagram should be displayed in the center of the page.
4] Choice of diagram: the choice of a particular type of diagram has to be made depending on nature of data magnitude of observations and type of people who use the diagram.
5] Index: an index illustrating different types of shades, colors, lines, designs etc used for presenting the data should be provided for easy Understanding of the diagram by the user.

Types of diagrams:
Some of the most commonly used diagrams are explained hear under:
One dimensional diagram:
One-dimensional diagram means the construction of a diagram on the basis of one dimension.

Line diagram:
A line diagram involves drawing multiple vertical lines. Each vertical line represents a particular frequency. The variety (x) values are presented on X axis and the corresponding frequencies for each value of x are presented on y axis. Both the axis is drawn as per a suitable scale.

Bar diagram: Bar diagrams are one of the most commonly used diagrams for presenting data. This type of diagram is generally constructed when the data represent different values over a period or different situations. They are largely used by business and government to present economy related data

such as industrial / agriculture production inflation exports. Fore reserves etc,
Bar diagrams are used specifically for categories data or series. They consist of a group of equidistant rectangles. One for each group or category of data in which the value of magnitudes are represented by the length or height of the rectangles.

Simple bar diagrams:
It is used for comparative study of two or more aspects of a single variable or single category of data. In this simple bar diagrams, single variable or simple classification or single category of data is represented.

Sub – divided or component bar diagram:
Simple bar diagrams are useful when data is homogenous and only one particular characteristic is being studied. However, if multiple aspects/ factors contributing to that one characteristic are to be studied, we will need to take the help of sub-divided bar diagram or component bar diagram.

Percentage bar diagram:
If subdivided bar diagrams are presented basis i.e., each component as a percentage of the whole, it is said to be percentage bar diagram also. If the component diagram is presented in percentage, it is termed as percentage bar diagram.

Multiple bar diagram:
In simple bar diagram, if is difficult to depict two or more variable. Thus, if two or more sets of data are to be presented simultaneously, multiple bar diagrams are used .In this diagram, two or more bars, which either represent different variables, or various components of the same variable are constructed adjoining each other. Multiple bar diagrams are used if two or more sets of inter – related phenomena or variables are to the presented graphically .this is done by drawing a set of adjacent bars, each representing a different variable.

Deviation bar diagram:
It is used for presentation of net quantities, which can be positive or negative. Net positive figures are presented by bars above the base line while negative figures are shown below the baseline. Deviation diagrams are

used to present the net quantities, when two sets of data are given and their net balance can be positive or negative.

Duo-directional diagrams:
Due-directional diagrams, as the name suggests are drawn in 2 directions. These diagrams are used to present an aggregate result of different and opposite components of the same phenomenon.

Broken bar diagram: it is used for presentation of data, which contain wide variations in values. Sometimes, the range of data is very wide. The data may contain very large observations along with small observation.

Pie diagram:
It enables us to show the breakup of given total into various component parts. It is so called because the entire looks like a pie and the components resemble slices cut from the pie. In order to construct a pie diagram equate the total of all values to 360 deg by drawing a circle with a radius of any length. For each component, find the number of degrees that can be assigned to it, if the total is 360 degrees.

Two – dimensional diagrams:
In two dimensional diagrams, both length and width of the observations are represented by bars. The area of a two – dimensional diagram is equal to the product of its length and width. So, two – dimensional diagrams are also called area diagrams. The various types of two – dimensional diagrams are briefly explained below:

a) Rectangular diagram :

A rectangle diagram is a modified form of bar diagram. It provides more detailed information than is furnished by a diagram. The length of the rectangle represents one aspect of data while its width would represent another aspect of the data provided.

Sub – divided rectangle diagram:
As in case of a bar diagram, a rectangle can also be subdivided or percentage sub- divided to represent multiple characteristics at the same time. The procedure for preparing a sub-dividing or percentage rectangle (two dimensional) is the same as in case of sub-divided/ percentage bar diagram.

b) Square diagram:

A square diagram also is a two dimensional diagram representing area. Square diagrams are specifically and if we wish to compare values that differ significantly from one other. c) Circle diagram:
Circle diagram is an alternative to square diagram. It is also used to present the values differing widely in their magnitudes.
Three – dimensional diagrams:
Three – dimensional diagrams are diagrams in which three dimensions, namely, length, breadth and height are taken into account.

Pictograms:
Pictures are more attractive and appealing to the eye and have a lasting impression on the mind. Pictograms are the technique of presenting statistical data through appropriate pictures.

Cartograms:
Cartograms are used to present data pertaining to geographical regions with the help of maps.
Choice of diagram:
Visual presentation of a given set of data is not easy and requires great skill, intelligence and expertise.

Graphs:
A student of economics studying the relationship between demand and price (law of demand) observe the quantity of mangoes being demanded by a housewife at various prices.

Technique of construction of graphs:
Graphs are drawn on a special type of paper known as graphs which has a fine network of horizontal and vertical lines for smaller parts of the same. In an each division of a centimeter and thin lines for smaller parts of the same.

General rules for graphing:
Apart from the various aspects discussed under general rules for constructing diagrams, the following points need to be kept in mind.
1] Neatness:

One of the basic requirements of a graph is to present data in a manner that is visually pleasing.

2] Title:

Every graph must have a clear and comprehensive title so that the user of the graph understands what is being presented by means of the graph.

3] Footnote:

Any explanation that is necessary to explain and is not obvious needs to be explained with the help of a footnote.

4] Structural framework:

The position of the axes should be so chosen that the graph looks proportionate and attractive.

Types of graphs:

Graphs can be broadly classified under two types:

Time series graphs or historigrams:

Time series graphs have time as an independent variable on the X-axis (horizontal axis) and dependent variables on the Y-axis (vertical axis), two –or more variables dependent on time can be conveniently presented on the same graph provided the magnitudes do not vary a great deal.

Time series graphs can be further classified into:

Horizontal line graphs:

These are the simplest of time series graphs. In case of horizontal lines graphs, separate lines represent the variables with reference to the time.

Band graphs:

A brand graph is prepared when a variable has several components. Hence band graph is also known as component graph.

Silhouette or net balance graph:

This graph is used when two different but relative variable need to be presented with clear emphasis on the difference between the two. In this graph, the net balance is highlighted with the help of distinct shades or such other depiction.

Range graph:

Range graph is used to depict the variation in data with reference to each period. The highest, the lowest and the average values of the variable being studied are plotted on the graph paper.

Graphs of frequency distribution:

Frequency graphs are designed to reveal clearly the characteristics features of a frequency data. In a frequency graph the size or the value of the item is presented on the horizontal axis and the frequency or the number of items on the vertical axis.

Vertical line graph:

A vertical line graph is used in case of discrete series. Wherein the phenomenon whose frequency is being studied is listed on the X axis and the frequency of the phenomenon is depicted in terms of the height of a straight line parallel to the Y axis.

Histogram:

A histogram is a graphical method of presenting data. Where the observations are located on a horizontal axis and the frequency of those observations is depicted along the vertical axis.

Points to note:

1] Histogram can be drawn only if the frequency distribution is continuous. In case the given series is an inclusive series, we can eliminate the gap between the limits of two adjacent classes by taking the mid-values of the class limits of two adjacent classes as the common limit of both the classes. In other words, we convert the series into an Exclusive series.

2] Sometimes, only the mid-values of different classes are given, in such a case the given distribution is converted into continuous classes by ascertaining the upper and lower limits of the various classes under the assumption that the frequency is uniformly distributed throughout the class intervals.

3] Histograms can also be used to represent discrete frequency distribution. This is done by considering the given values of the variable as the mid-points of continuous classes and then proceeds as explained above.

Difference between histogram and bar diagram:

i) A histogram is two dimensional (area) diagram where both the width (base) and the length (height of the rectangle) are important whereas bar diagram is one dimensional diagram in which only length (height of the bar) matters while width is arbitrary.

ii) In a histogram, the bars (rectangle) are adjacent to each other whereas in bar diagram proper spacing is given between different bars.

Frequency polygon:

Polygon is a figure with multiple, usually more than four side. Frequency polygon is graphical representation of frequency distribution in the form of a curve superimposed on a histogram.

Without histogram:

A frequency polygon is drawn on the presumption that the frequencies in a class interval are evenly distributed throughout the class and hence, their mid-points are representative in actual construction.

Frequency curve:

Frequency curve is a smooth free hand curve drawn through the vertices of frequency polygon.

Types of frequency curves:

There are different types of frequency curves:

1] Symmetrical bell-shaped:

The curve is bell shaped and can be bisected at the center to result in two equal sides mirroring each other.

2] Asymmetrical bell shaped:

The curve is bell shaped but cannot be bisected at the center. The peak is either towards the left or right of the center, if the peak is towards the left, if is termed as positively skewed. If the peak is towards the right, it is termed as negatively skewed.

Ogive:

Ogive is a graphic representation of a cumulative frequency of a distribution. It is also known as cumulative frequency polygon. It is pronounced as ogive.

Less than ogive:

The less than ogive like an elongated S which starts from the lowest class boundary on the horizontal axis and gradually rises upward, ending at the highest class boundary corresponding to the total frequency of the distribution.

More than ogive:

The more than ogive has the appearance of an
Elongated S turned upside down. It starts from the lowest class boundary on the horizontal axis and gradually slopes downward, ending at the highest class boundary corresponding to the total frequency of the distribution.

Lorenz curve:

Lorenz curve is a special type of graph designed to show how much a certain distribution varies from a completely uniform distribution.

Interpretation:

The Lorenz curve is a graphic method of indicating whether a certain quantity is equally distributed throughout the population.

Difference between diagrams and graphs

Diagrams:

Plain paper is used

Diagrams help in comparison of variables and not for study of mathematical relationship between them

Data is presented by devices such as bars, rectangles, squares, circles, cubes, etc

Graphs:

Graph paper is used

Graphs help in study of mathematical relationship between the variable

Points or lines of different kinds such as dots, dashes, dot-dash, etc, are used to present the data.

UNIT – III

MEASURES OF CENTRAL TENDENCY

Meaning and introduction:

We have learnt in previous chapters as how to large masses of primary data can be classified, tabulated and presented to enable the user to analyze the data. While this data is very useful for the statistician, the common man would look for a number that is representative of the entire data. In the words of prof. R.A. fisher ,the inherent inability of the human mind to grasp in its entirely a large body of numerical data compels us to seek relatively few constants that will adequately describe the data.

Definition:

The terms average of measures of central tendency have been defined by various authors. Some of the definitions are given below.

According to Simpson and Kafka a measure of central tendency is a typical value around which other figures congregate. It is a single value that represents a whole series.

In the words of Ya Lun Chou, an average is a typical value in the sense that it is sometimes employed to represent all the individual values in a series values in a series of a variable.

Objectives / functions / uses of average:

Averages serve a very useful purpose in statistics.

Following are the objectives of averaging data:

1] Represent the characteristics of the entire mass of data:

An average reduces a complex mass of data into a single typical figure to enable the user to get a bird s eye view about the characteristics of the phenomenon under study. 2] To Facilitate Comparison:

An average reduces a mass of complex data into one single figure. This makes it possible for the user to make comparisons between two or more sets of data and draw conclusions the characteristics of the separate sets of data.

3] Describe data:

Average serves the objective of describing the main characteristics of the underlying data in a simple and brief manner.

Limitations of averages:

Averages are very useful statistical tools, but they also suffer from certain limitations.

The following are the limitations of averages:

1) Misleading conclusions:

Average is a single numerical figure representing the characteristics of a given distribution. This number is vulnerable to errors in interpretation and can lead to misleading conclusions.

Choice of average:

There are different types of averages. Different types of averages are suitable for different objectives. The utility of average depends on a proper and judicious choice of the average.

Incomplete picture:

An average does not provide the complete picture of a distribution.

Characteristics of a good average:

The impact of some of the aforementioned limitations is not significant if the average being used has certain desirable characteristics.

The following are the essentials of a good average:

1] It should be rigidly defined: An average should be rigidly defined so that there is no scope for confusion or manipulation. 2] It should be easy to calculate and simple to follow:

Calculation of an average should be simple to understand. It should be easy to calculate, preferably without the help of calculators.

Types of averages:

Averages are of the following types: A)

Mathematical averages:

I) arithmetic average or mean

II) geometric mean

III) harmonic mean B) Averages of position:

i) median
ii) mode

Of the above, arithmetic mean, median and mode are most widely used, followed by geometric mean and harmonic mean. Let us understand each of these averages.

Merits of arithmetic average

1) It is rigidly defined. Hence, different interpretations by different persons are not possible.
2) It is easy to understand and easy to calculate. In most of the series it is determinate and its value is definite.
3) It takes all values into consideration. Thus, it is more representative.
4) It can be subjected to further mathematical treatment. The properties of arithmetic mean are separately explained elsewhere in the chapter.

Demerits of arithmetic average:

While arithmetic mean satisfies most of the conditions of an ideal average, it suffers from certain drawbacks. Some of the demerits or limitations of arithmetic mean are listed below:

It cannot be determined by inspection

It cannot be located graphically

It cannot be used in the study of qualitative phenomena.

Meaning of median:

Median, in simple terms, means middle. In statistics, median represents the middle point in a series. Median divides the observations in two equal parts, in such a way that the number of observations smaller than median is equal to number of observations greater than it.

Definition of median:

According to prof. Horace secrist, median of a series is the value of the item, actual or estimated, when a series is arranged in order of magnitude which divides the distribution into two parts.

Merits of median:

1] It is rigidly defined

2] It is easy to understand and easy to calculate

3] It is not influenced much by items on the extremes.
4] It can be computed for distributions which have open end classes.

Demerits of median:
1] Median is not exact in many cases.
2] Calculations of median do not consider all the items of the series. Thus it is not fully representative of the entire data.
3] Median is significantly impacted by fluctuations of sampling. Hence, it is less reliable.

Mode:
While looking for a shirt is one of the retail outlets, you will probably find a number of shirts available for different sizes. However, the number of shirts as well as the variety will be more for some sizes (size 40for example) when compared to other sizes (size 44 for example). How does the shirt manufacturer or the store owner decide on which sizes to store more? This decision can be taken very easily by applying the concept of mode.

Definition of mode:
According to Croxton and cowden, the mode of distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded as the most typical of a series of values. In the words of A.M. Tuttle, mode is the value which has the greatest frequency density in its immediate neighborhood.
Merits of mode:

1) Mode is easy to understand and calculate
2) It is not influence much by items on the extremes
3) It can be located even if the class- intervals are of unequal magnitudes provided the modal class and the preceding and succeeding it is of the same magnitude.
4) It can be computed for distributions which have open end classes
5) Mode is not an isolated value like the median. It is the term that that occurs most in the series.

Drawbacks of mode

1] Calculation of mode does not consider all the items of the series. Thus, it is not fully representative of the entire data.
2] It is not rigidly defined
3] It is not capable of further mathematical treatment
4] Mode is sometimes indeterminate. They may be 2 (bi-modal) or more (multimodal) values.

Geometric mean:

Geometric mean is the root of the product of items of a series. If there are 2 numbers, say a and b, the geometric mean of the two numbers is the square root of the product of the 2 numbers.

Merits of geometric mean:

1] It is rigidly defined. Hence, different interpretations by different persons are not possible
2] It takes all values into consideration. Thus, it is more representative.
3] It can be subjected to further mathematical treatment. The properties of geometric mean have been separately explained

Drawbacks of geometric mean:

1) It is neither simple to understand nor easy to calculate
2) It cannot be determined by inspection
3) It cannot be located graphically
4) It cannot be used in the study of qualitative phenomena.

Unit – IV A

Measures of dispersion Meaning and definition:

An average may not provide us insight into the characteristics of any data unless we understand the manner in which the individual data items scatter around the average. This phenomenon is called dispersion. It measures the extent to which the items vary from some central value.

According to Bowley, dispersion is the measure of the variation of items.

Objectives of dispersion:

The main objectives of dispersion are

1) To gain a better understanding of a given data series by understanding the average variation and the dispersion of values on either side of the measure of central tendency
2) To establish whether a measure of central tendency of a data series is truly representatives of the series
3) To understand the range of values that the given data series can take

Properties of a good measure of dispersion:

A good measure of dispersion must have the following characteristics or properties: 1. It must be easy to calculate

2. It should be simple enough to understand the message the measure is trying to convey

3. It must be rigidly defined. While every value in the series might have an impact on the measure, it must not be capable of completely changing the measure such that the conclusions are totally different.

Measures of dispersion:

There are different measures of dispersion. These measures can be classified into absolute measures and relative measures.

Absolute measures

1) The range
2) Inter quartile range and quartile deviation
3) The mean deviation or average deviation
4) The standard deviation and variance
5) The Lorenz curve

Relative measures:
1] Coefficient of range
2] Coefficient of quartile deviation
3] Coefficient of mean deviation and variance

4] Coefficient of variation Range:
Range is the difference between the values of the extreme items of a series. It is the difference between the values of the largest item and the value of the smallest items in the distribution.

Merits of range:
1] It is the simplest to calculate of all measures
2] It is very easy to understand
3] It gives a high level or broad picture of the data at one glance

Demerits of range:
1] It is based only on the smallest and largest values of a series and hence is unduly impacted by extreme values. In other words, it is not based on all the terms and hence, not rigidly defined 2] Range is impacted by sample size and composition of the sample
3] Range cannot be calculated for open ended distributions

Quartile deviation:
Quartile deviation shows the average amount by which the two quartiles differ from mean Merits of quartile deviation:
1] It is the simple to calculate and very easy to understand
2] It is not impacted by extreme values
3] It can be computed for open ended distributions and for data containing unequal classes

Demerits of quartile deviation:

1] It is impacted by sample size and composition of the sample
2] It is not amenable to algebraic or statistical treatment

Mean deviation:
Mean deviation of a series is the arithmetic average of the deviations of various items from a measure of central tendency (mean, median or mode)

Merits of mean deviation:
1] It is rigidly defined
2] It is not least impacted by sampling fluctuations
3] It takes into account every single value in the series 4] Mean deviation from median is least impacted due to extreme values

Demerits of mean deviation:
1] It is relatively difficult to compute
2] It is not amenable to further algebraic or statistical treatment

Standard deviation:
Standard deviation is the square root of the arithmetic average of the square of the deviations measured from mean

Merits of standard deviation:
1] It is rigidly defined
2] It takes into account every single value in the series
3] It is amenable to further algebraic or statistical treatment

Demerits of standard deviation:
1] It is relatively difficult to compare
2] It is calculated with only arithmetic mean as the average. Standard deviation from other averages such as median is not an effective measure of dispersion

UNIT – IV B

Skewness and kurtosis Meaning:

Skewness means lack of symmetry in a frequency distribution. Symmetry is implied when data values are distributed in the same way above and below a middle value.

Definition:

According to croxton when a series is not symmetrical it is said to be asymmetrical or skewed. Skewness is defined by Spiegel as the degree of asymmetry or departure from symmetry of a distribution.

Types of skewness:

In a distribution, skewness is present if

1] The values of mean, median and mode do not coincide 2] The sum of the positive deviations from the median is not equal to the sum of the negative deviations

3] Quartiles are not equidistant from the median

4] Data when plotted on a graph paper will not give normal bellshaped curve

Difference between dispersion and skewness:

1] Dispersion is concerned with measuring the amount of variation in a series. Skewness is concerned with direction of variation or the departure from symmetry

2] Dispersion gives us the extent and direction in which the distribution differs from symmetry

3] Dispersion tells us about the composition of the series. Skewness tells us about the shape of the series.

Moments and kurtosis Meaning:

The word kurtosis is derived from the Greek word kyrtos which means curved.kurtosis describes the shape of a probability distribution. Kurtosis tells us whether the distribution, when plotted on a graph, would give us a normal curve, a curve that is more flat than the normal curve or a curve that is more peaked than the normal curve.

Definition:
According to Croxton and cowden a measure of kurtosis indicates the degree to which a curve of a frequency distribution is peaked or flat topped.

Comparison among dispersion, skewness and kurtosis:
Dispersion, skewness and kurtosis are different characteristics of frequency distribution. Dispersion studies the scatter of the items round a central value or among themselves.

UNIT - V

Correlation

Meaning and definition:
Correlation means the relationship between two variables where, with the changes in the values of one variable, the values of other variable also change. According to croxten and crowden , 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 ya lun Chou, correlation analysis attempts to determine the degree of relationship between variables.

Types of correlation:

1] Positive or negative correlation:

Correlation can be either positive or negative. When the values of two variables move in the same direction so that an increase in the value of one variable is associated with an increase in the value of one variable is associated with an increase in the value of other variable also, and a decrease in the value of one variable is associated with the decrease in the value of other variable also, correlation is said to be positive.

2] Linear and non linear correlation:

Correlation between two variables is said to be linear if the change in one variable in response to change in another variable is proportionate.

3] Simple and multiple correlations:

When the relationship between two variables is studied, we are trying to establish a simple correlation. 4] Partial and total correlation:

When there are more than 2 variables but the relationship between two variables is studies assuming that the other variables are constant, such correlation is called partial correlation.

Methods of studying correlation:

The following methods are usually adopted to study correlation between variables.

1] Scatters diagram method
2] Correlation graph
3] Karl pearson s coefficient of correlation
4] Coefficient of correlation by rank differences
5] Coefficient of concurrent deviation

Properties of karl pearson s coefficient of correlation:

1] It is based on arithmetic mean and standard deviation .
2] It lies between – 1
3] It measures both direction as well as degree of change
4] It is independent of change in scale. In other words, if a constant amount is added/subtracted from all values of a variable, the value of r does not change.

5] It is independent of change in origin. Thus, if a constant amount is multiplied with or divides all values of a variable, r does not change.

Merits of Karl Pearson s coefficient of correlation:
1] It takes into account all items of the variable (s)
2] It is a numerical measure and hence more objective 3] It measures both direction as well as degree of change 4] It facilitates comparisons between two series.

Demerits of Karl Pearson s coefficient of correlation:
1] It is not easy to calculate as complex formulae are involved 2] It is more time consuming compared to methods such as rank correlation
3] It assumes a linear relationship between the two variables which may not be correct.

Properties of spearman s rank correlation:
1] It is based on subjective ranking of variables
2] It lies between – 1
3] It measures both direction as well as degree of change.

Merits of spearman s rank correlation:
1] It is easy to understand and calculate
2] It is not impacted by extreme values
3] It is a numerical measure and provides objectively to subjective ranking

Demerits of spearman s rank correlation:
1] It cannot be applied to grouped data
2] It lacks the precision of karl pearson s coefficient of correlation
3] All the information concerning the variables is not used
4] The computation becomes complicated as the number of observations increase.

**********END**********