BUSINESS STATISTICS

Statistics- Meaning:

The word “Statistics” has been derive from the Latin word “Status” or Italian word “Statista” or German word “Statistika”. Each of these words means Political State. Initially, Statistics was used to collect the information of the people of the state about their income, health, illiteracy and wealth etc. But now a day, Statistics has become an important subject having useful application in various fields in day to day life. Statistics.

Characteristics of Statistics

Main characteristics of Statistics in terms of numerical data are as follows:

(1) Aggregate of Facts – A single number does not constitute Statistics. We can not draw any conclusion from single number. We can draw any conclusion by the aggregate number of facts. For example, if it is stated that there are 1,000 students in our college then it has no significance. But if it is stated that there are 300 students in arts, 400 students in commerce and 300 in science in our college. It makes statistical sense as this data convey statistical information.

(2) Numerically Expressed - Statistics are expressed in terms of numbers. Qualitative aspects like small or big, rich or poor etc. are not statistics. For instance if we say that Irfan Pathan is tall Sachin is short then this statement has no statistical sense. However if it is stated that height of Irfan Pathan is 6 ft and 2 inch and the height of Sachin is 5 ft and 4 inch then these numerical will be called Statistics.

(3) Affected by Multiplicity of Causes – Statistics are not affected by any single factor but it is affected by many factors. For instance 30% rise in prices may have been due to several causes like reduction in supply, increase in demand, shortage of power, rise in wages, rise in taxes, etc.

(4) Reasonable Accuracy - A reasonable degree of accuracy must be kept in view while collecting statistical data. This accuracy depends on the purpose of investigation, its nature, size and available resources.

(5) Pre-determined Purpose Statistics are collected with some pre-determined objective. Any information collected without any definite purpose will only be a numerical value and not Statistics. If data pertaining to the farmers of a village is collected, there must be some pre-determined objective. Whether the statistics are collected for the purpose of

knowing their economic position or distribution of land among them or their total population. All these objectives must be pre – determined.

(6) Collected in a Systematic Manner – Statistics should be collected in a systematic manner. Before collecting the data, a plan must be prepared. No conclusion can be drawn from data collected in haphazard manner. For instance, data regarding the marks secured by the students of a college without any reference to the class, subject, examination, or maximum marks, etc will lead no conclusion.a

IMPORTANCE OF STATISTICS

In recent days, we hear talking about statistics from a common person to highly qualified person. It only show that how statistics has been intimately connected with wide range of activities in daily life. They realize that work in their fields require some understanding of statistics. It indicates the importance of the statistics. A.L.Bowley says, "Knowledge of statistics is like knowledge of foreign language or of algebra. It may prove of use at any time under any circumstances

1. **Importance to the State or Government;**

In modern era, the role of state has increased and various governments of the world also take care of the welfare of its people. Therefore, these governments require much greater information in the form of numerical figures. Statistics are extensively used as a basis for government plans and policies. For example-5-years plans are framed by using reliable statistical data of different segments of life.

2. **Importance in Human Behavior;**

Statistical methods viz., average, correlation etc. are closely related with human activities and behavior. For example-when a layman wishes to purchase some article, he first enquiries about its price at different shops in the market. In other words, he collects data about the price of a particular article and aims at getting idea about the average of the prices and the range within which the price vary. Thus, it can be concluded that statistics play an important role in every aspect of human activities and behavior.

3. **Importance in Economics;**

Statistics is gaining an ever increasing importance in the field of economics. That is why Tugwell said, "The science of economics is becoming statistical in its method." Statistics and economics are so interrelated to each other that the new disciplines like econometrics and economic statistics have been developed. Inductive method of generalization used in economics, is also based on statistical principle. There are different segments of economics where statistics are used.

(**A) Consumption**- By the statistics of consumption we can find the way in which people in different group spend their income. The law of demand and elasticity of demand in the field of consumption are based on inductive or inferential statistics.

(B) Production- By the statistics of production supply is adjusted according to demand. We can find out the capital invested in different productive units and its output. The decision about what to produce, how much to produce, when to produce is based on facts analyzed statistically.

(C) Distribution- Statistics play a vital role in the field of distribution. We calculate the national income of a country by statistical methods and compare it with other countries. At every step we require the help of figures without them. It is difficult to move and draw inferences.

4. Importance in Planning; for the proper utilization of natural and manual resources, statistics play a vital role. Planning is indispensable for achieving faster rate of growth through the best use of a nation's resources. Sometimes said that, "Planning without statistics is a ship without rudder and compass." For example- In India, a number of organizations like national sample survey organization(N.S.S.O.), central statistical organization (C.S.O.) are established to provide all types of information.

5. Importance in Business: The use of statistical methods in the solution of business problems dates almost exclusively to the 20th century. Or now days no business, large or small, public or private, can prosper without the help of statistics. Statistics provides necessary techniques to a businessman for the formulation of various policies and planning with regard to his business. Such as-

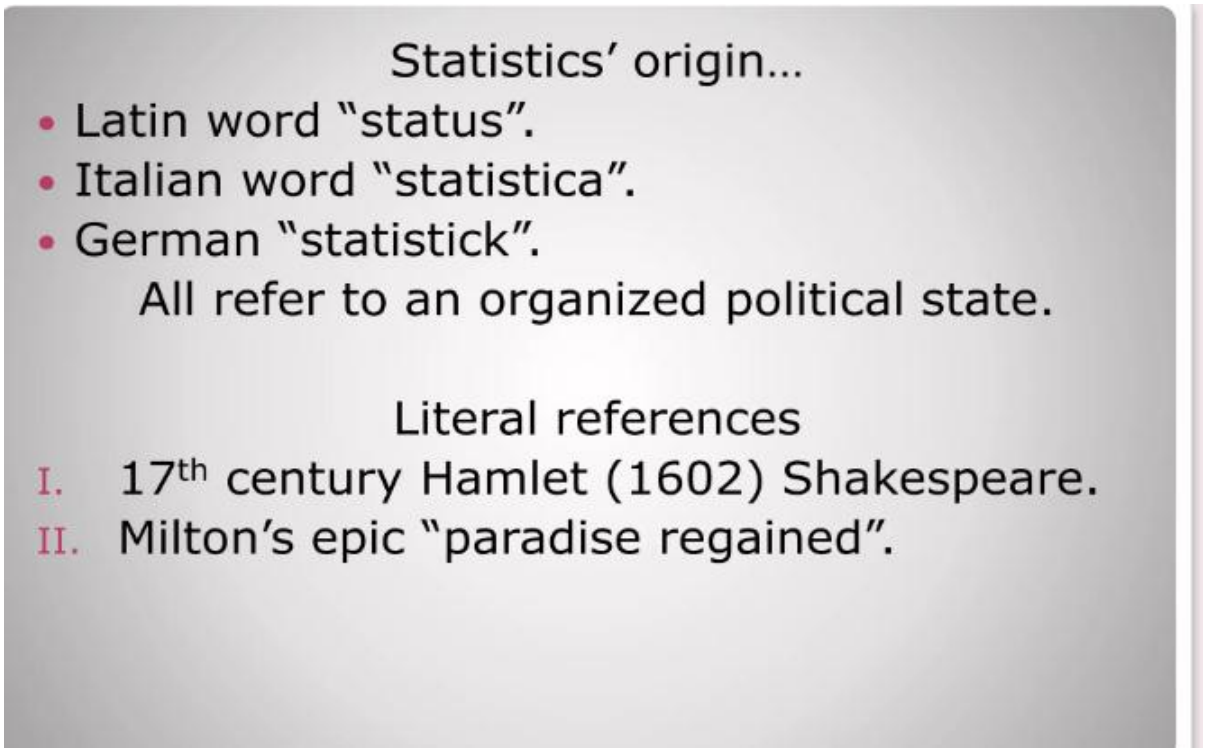
(A) Marketing- In the field of marketing, it is necessary first to find out what can be sold and them to evolve a suitable strategy so that goods reach the ultimate consumer. A skillful analysis of data on population, purchasing power, habits of people, competition, transportation cost etc. should precede any attempt to establish a new market.

(B) Quality Control- To earn the better price in a competitive market, it is necessary to watch the quality of the product. Statistical techniques can also be used to control the quality of the product manufactured by a firm. Such as - Showing the control chart.

(C) Banking and Insurance Companies- banks use statistical techniques to take decisions regarding the average amount of cash needed each day to meet the requirements of day to day transactions. Various policies of investment and sanction of loans are also based on the analysis provided by statistics.

(D) Accounts writing and Auditing- Every business firm keeps accounts of its revenue and expenditure. Statistical methods are also employed in accounting. In particular, the auditing function makes frequent application of statistical sampling and estimation procedures and the cost account uses regression analysis.

(E) Research and Development- Many business organizations have their own research and development department which are responsible for collection of such data. These departments also prepare charts groups and other statistical analysis for the purpose.



Modern definition by scholar:

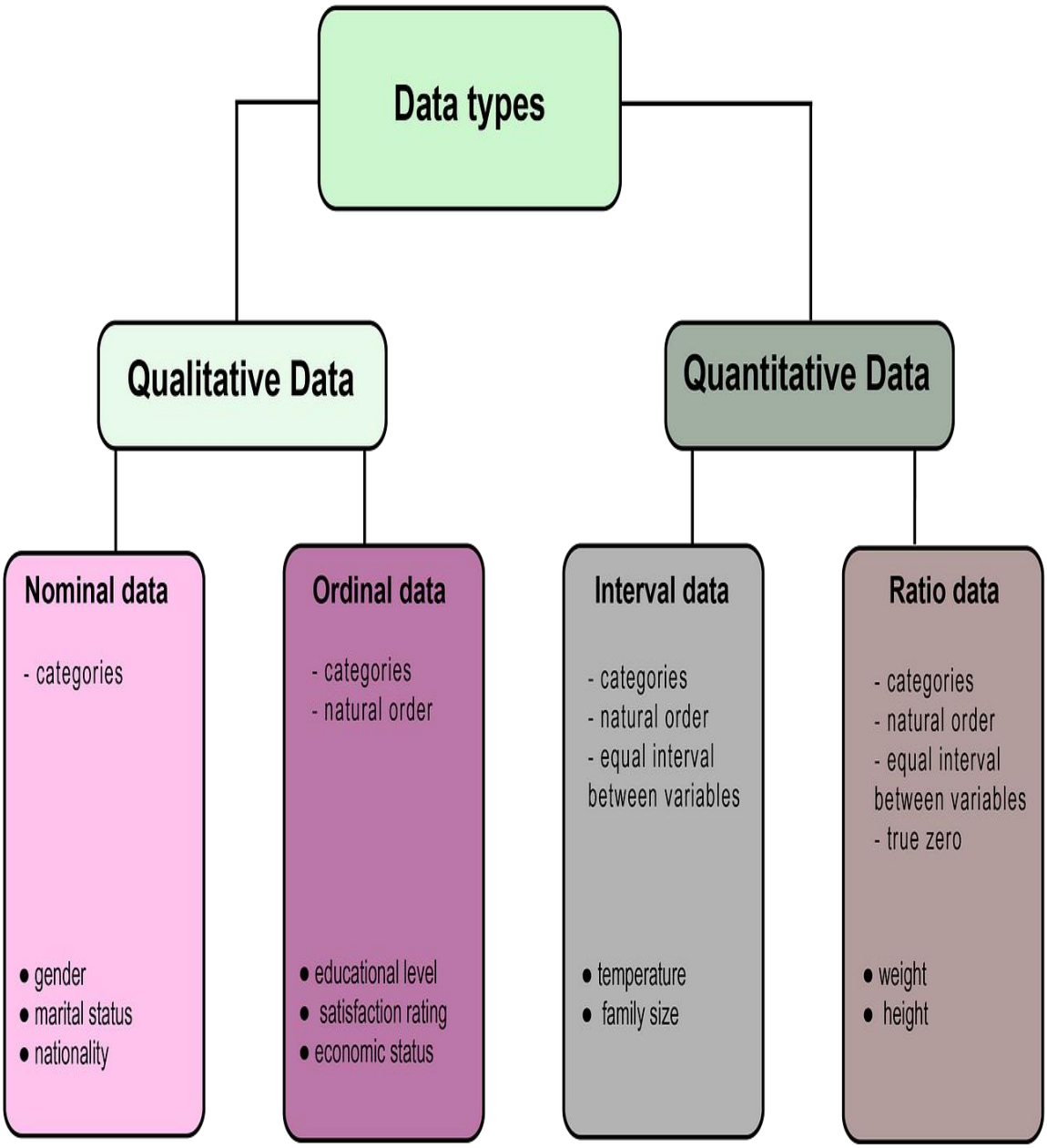
W.I.king states...

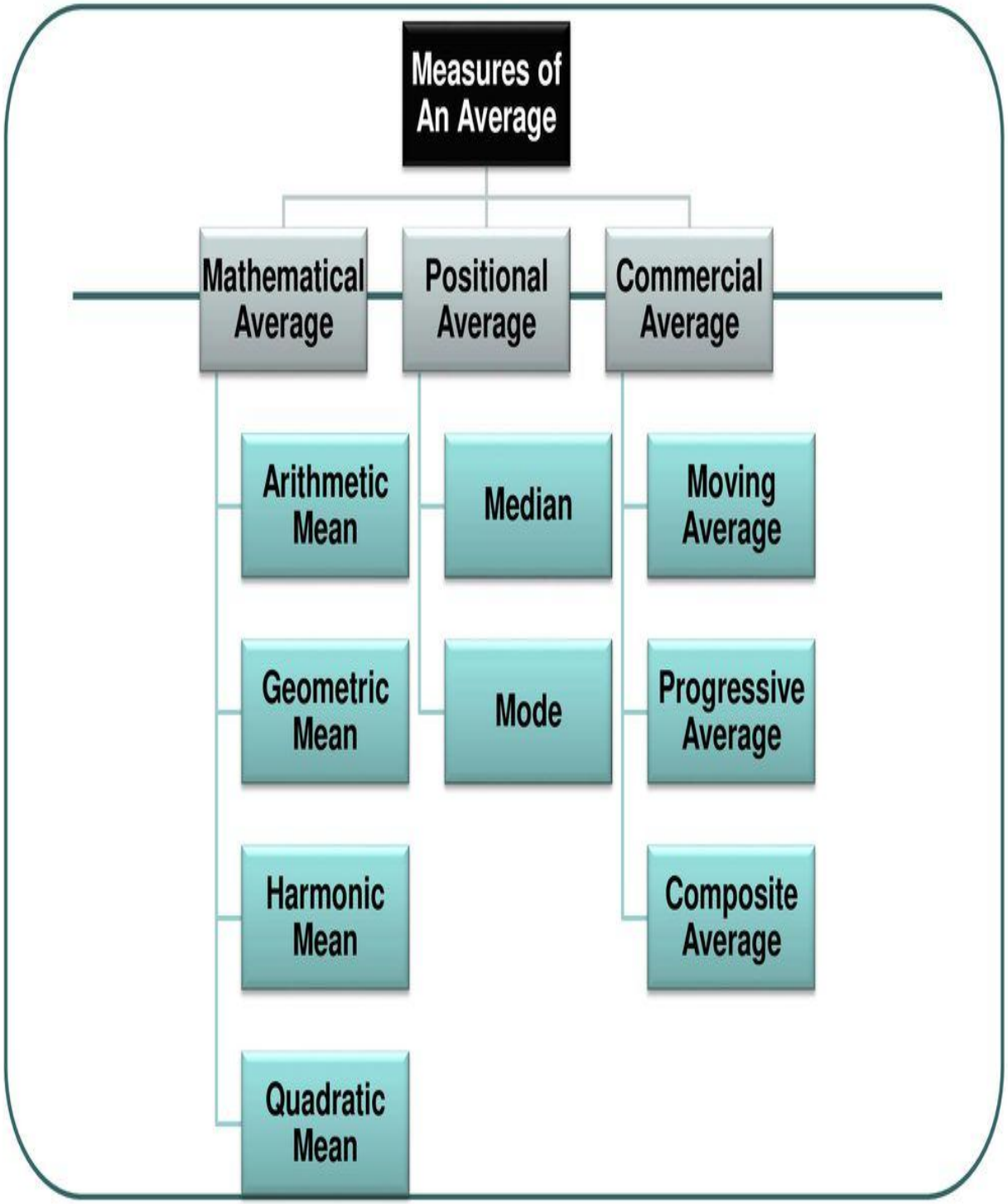
The science of statistics is the method of judging collective, natural or social phenomena from the results obtained by the analysis of an enumeration or collection of estimates.

Characteristics/features :

- ✓ Its basically an aggregate of facts that are numerically expressed with perfect logic which make it short and easier to understand. The numbers rather than big explanation makes it more interesting.
- ✓ Statistics are affected to a marked extent by multiplicity of causes. Since stats is made up from facts arranged, all the sourced factors affect its accuracy and reliability.

- ✓ Collected in standard accuracy which is expected in a systematic manner for a predetermined purpose.
- ✓ Statistics are capable of being placed in relation to each other. i.e. since statistical data are mostly collected for the purpose of comparison, the data ought to be homogenous and uniform.

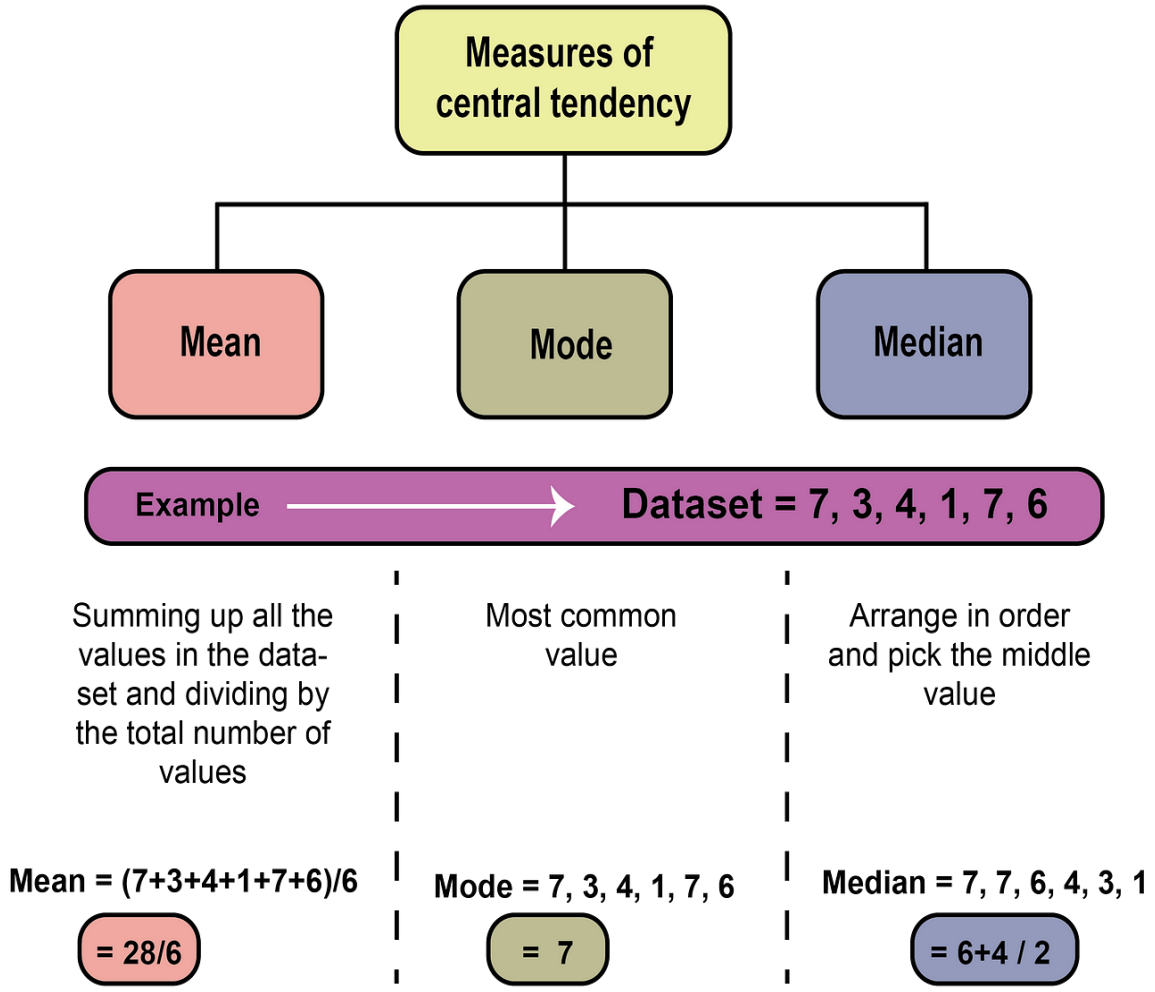

Data types
Qualitative Data
Quantitative Data
Nominal data
- categories
• gender
• marital status
• nationality
Ordinal data
- categories
- natural order
• educational level
• satisfaction rating
• economic status
Interval data
- categories
- natural order
- equal interval between variables
• temperature
• family size
Ratio data
- categories
- natural order
- equal interval between variables
- true zero
• weight
• height


Measures of An Average
Mathematical Average
Positional Average
Commercial Average
Arithmetic Mean
Geometric Mean
Harmonic Mean
Quadratic Mean
Median
Mode
Moving Average
Progressive Average
Composite Average

Vocabulary Flashcards for Students

	the average of a set of numbers
	the middle number in a set of data
	the number which appears most often in a set of numbers
	the difference between the lowest and highest numbers in a data set

www.classroomcompulsion.blogspot.com



Correlation refers to a process for establishing the relationships between two variables. You learned a way to get a general idea about whether or not two variables are related, is to plot them on a "scatter plot". While there are many measures of association for variables which are measured at the ordinal or higher level of measurement, correlation is the most commonly used approach.

Correlation in Statistics

This section shows how to calculate and interpret correlation coefficients for ordinal and interval level scales. Methods of correlation summarize the relationship between two variables in a single number called the correlation coefficient. The correlation coefficient is usually represented using the symbol r, and it ranges from -1 to +1.

A correlation coefficient quite close to 0, but either positive or negative, implies little or no relationship between the two variables. A correlation coefficient close to plus 1 means a positive relationship between the two variables, with increases in one of the variables being associated with increases in the other variable.

A correlation coefficient close to -1 indicates a negative relationship between two variables, with an increase in one of the variables being associated with a decrease in the other variable. A correlation coefficient can be produced for ordinal, interval or ratio level variables, but has little meaning for variables which are measured on a scale which is no more than nominal.

For ordinal scales, the correlation coefficient can be calculated by using Spearman's rho. For interval or ratio level scales, the most commonly used correlation coefficient is Pearson's r, ordinarily referred to as simply the correlation coefficient.

Correlation Coefficient

The correlation coefficient, r, is a summary measure that describes the extent of the statistical relationship between two interval or ratio level variables. The correlation coefficient is scaled so that it is always between -1 and +1. When r is close to 0 this means that there is little relationship between the variables and the farther away from 0 r is, in either the positive or negative direction, the greater the relationship between the two variables.

The two variables are often given the symbols X and Y. In order to illustrate how the two variables are related, the values of X and Y are pictured by drawing the scatter diagram, graphing combinations of the two variables. The scatter diagram is given first, and then the method of determining Pearson's r is presented. From the following examples, relatively small sample sizes are given. Later, data from larger samples are given.

Scatter Diagram

A scatter diagram is a diagram that shows the values of two variables X and Y, along with the way in which these two variables relate to each other. The values of variable X are given along the horizontal axis, with the values of the variable Y given on the vertical axis.

Later, when the regression model is used, one of the variables is defined as an independent variable, and the other is defined as a dependent variable. In regression, the independent variable X is considered to have some effect or influence on the dependent variable Y. Correlation methods are symmetric with respect to the two variables, with no indication of

causation or direction of influence being part of the statistical consideration. A scatter diagram is given in the following example. The same example is later used to determine the correlation coefficient.

Types of Correlation

The scatter plot explains the correlation between the two attributes or variables. It represents how closely the two variables are connected. There can be three such situations to see the relation between the two variables –

- Positive Correlation – when the values of the two variables move in the same direction so that an increase/decrease in the value of one variable is followed by an increase/decrease in the value of the other variable.
- Negative Correlation – when the values of the two variables move in the opposite direction so that an increase/decrease in the value of one variable is followed by decrease/increase in the value of the other variable.
- No Correlation – when there is no linear dependence or no relation between the two variables.

Regression

Regression is a statistical method used in finance, investing, and other disciplines that attempts to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables).

Also called simple regression or ordinary least squares (OLS), linear regression is the most common form of this technique. Linear regression establishes the linear relationship between two variables based on a line of best fit. Linear regression is thus graphically depicted using a straight line with the slope defining how the change in one variable impacts a change in the other. The y-intercept of a linear regression relationship represents the value of one variable when the value of the other is zero. Non-linear regression models also exist, but are far more complex.

Regression analysis is a powerful tool for uncovering the associations between variables observed in data, but cannot easily indicate causation. It is used in several contexts in business, finance, and economics. For instance, it is used to help investment managers value assets and understand the relationships between factors such as commodity prices and the stocks of businesses dealing in those commodities.

Regression as a statistical technique should not be confused with the concept of regression to the mean (mean reversion).

LEAST SQUARES METHOD

The least squares method is a form of mathematical regression analysis used to determine the line of best fit for a set of data, providing a visual demonstration of the relationship between the data points. Each point of data represents the relationship between a known independent variable and an unknown dependent variable. This method is commonly used by statisticians and traders who want to identify trading opportunities and trends.

The least squares method is a form of regression analysis that provides the overall rationale for the placement of the line of best fit among the data points being studied. It begins with a set of data points using two variables, which are plotted on a graph along the x- and y-axis. Traders and analysts can use this as a tool to pinpoint bullish and bearish trends in the market along with potential trading opportunities.

The most common application of this method is sometimes referred to as linear or ordinary. It aims to create a straight line that minimizes the sum of squares of the errors generated by the results of the associated equations, such as the squared residuals resulting from differences in the observed value and the value anticipated based on that model.

For instance, an analyst may use the least squares method to generate a line of best fit that explains the potential relationship between independent and dependent variables. The line of best fit determined from the least squares method has an equation that highlights the relationship between the data points.

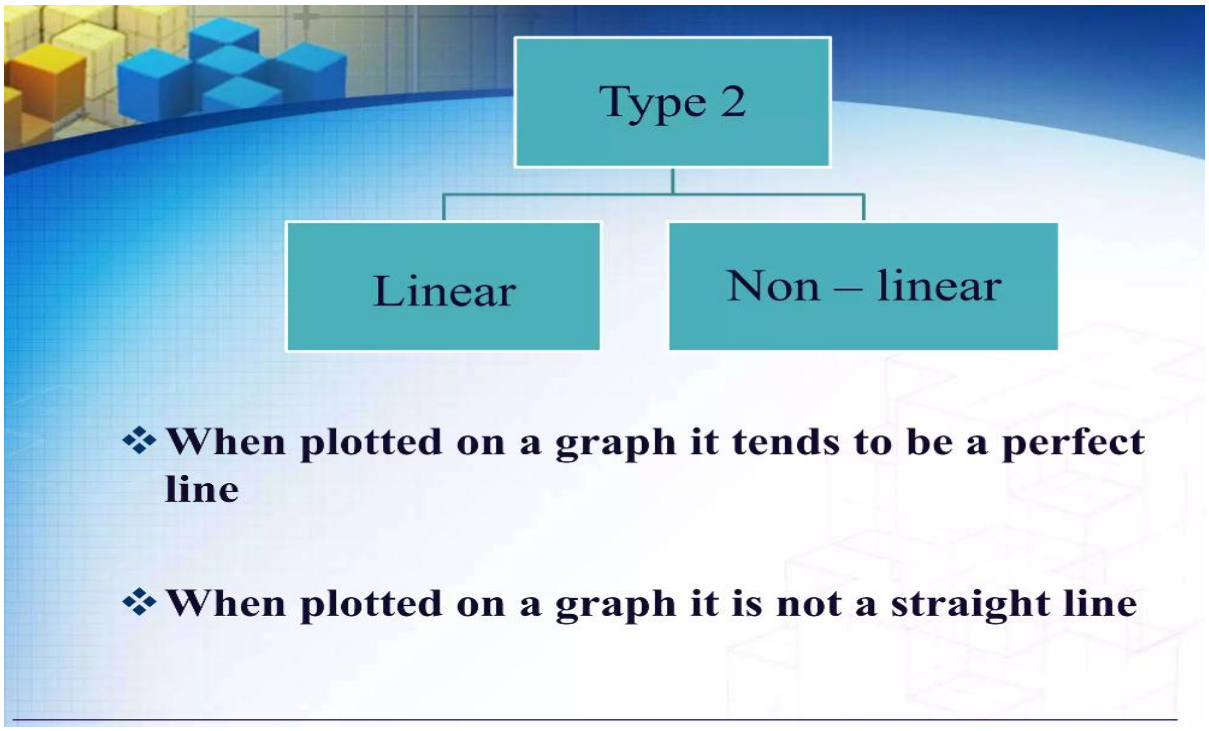
Introduction

- Correlation a LINEAR association between two random variables
- Correlation analysis show us how to determine both the nature and strength of relationship between two variables
- When variables are dependent on time correlation is applied
- Correlation lies between +1 to -1

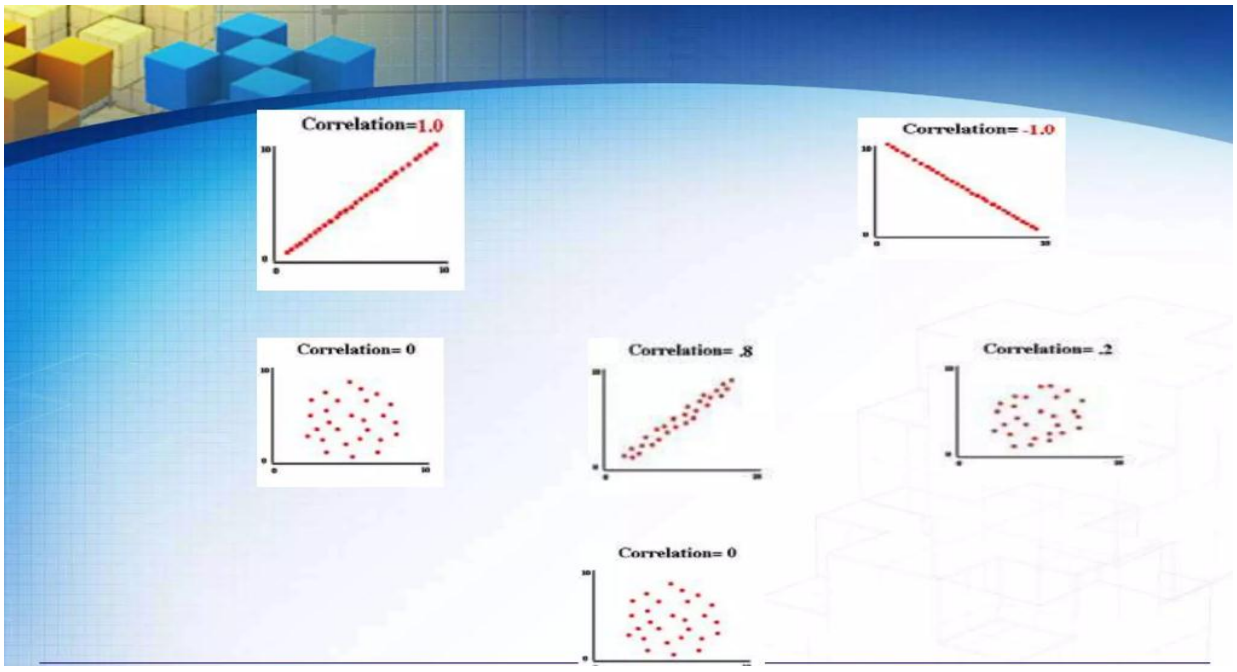
- A zero correlation indicates that there is no relationship between the variables
- A correlation of –1 indicates a perfect negative correlation
- A correlation of +1 indicates a perfect positive correlation

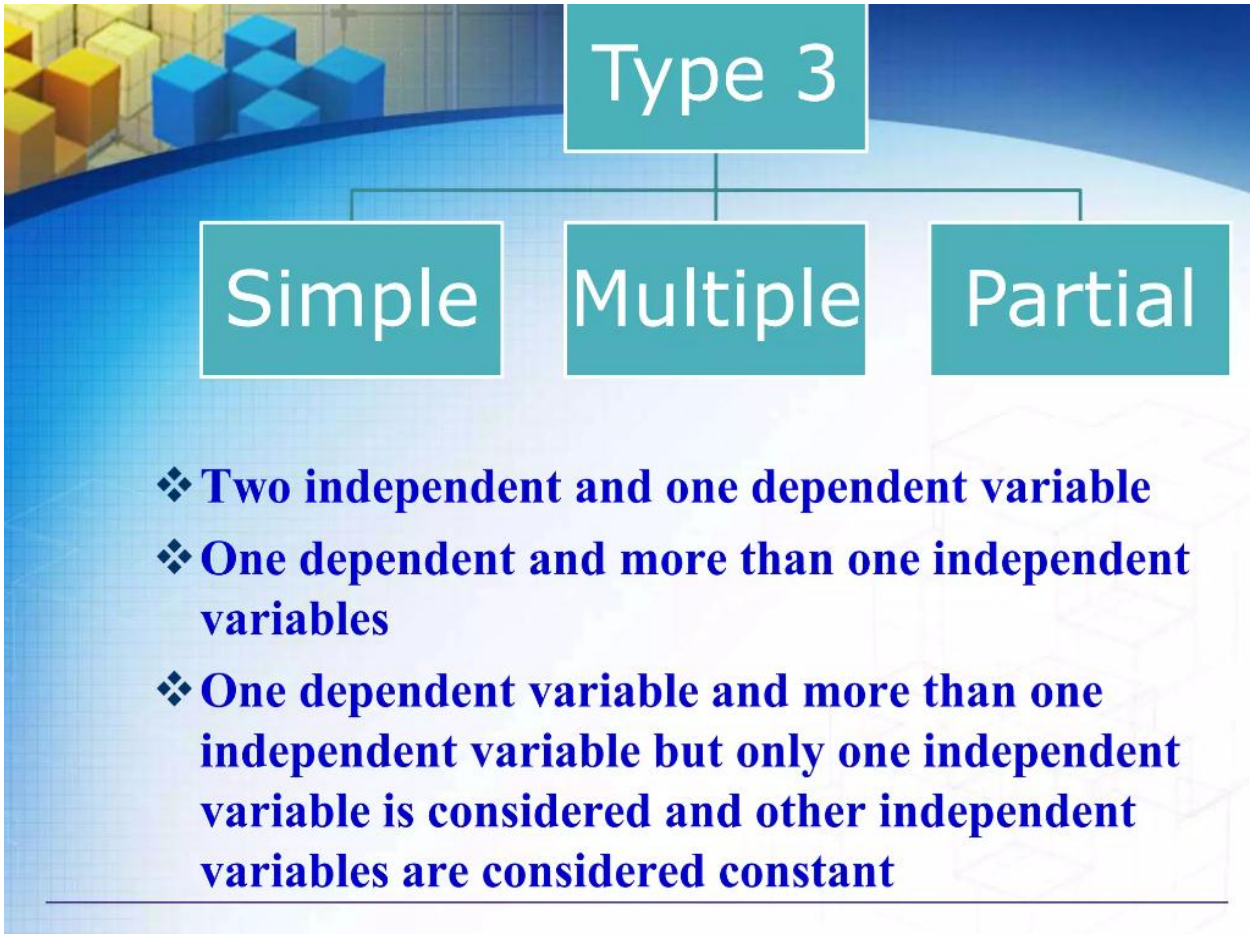
- A zero correlation indicates that there is no relationship between the variables
- A correlation of –1 indicates a perfect negative correlation
- A correlation of +1 indicates a perfect positive correlation

- If two related variables are such that when one increases (decreases), the other also increases (decreases).
- If two variables are such that when one increases (decreases), the other decreases (increases)
- If both the variables are independent



- **When plotted on a graph it tends to be a perfect line**
- **When plotted on a graph it is not a straight line**





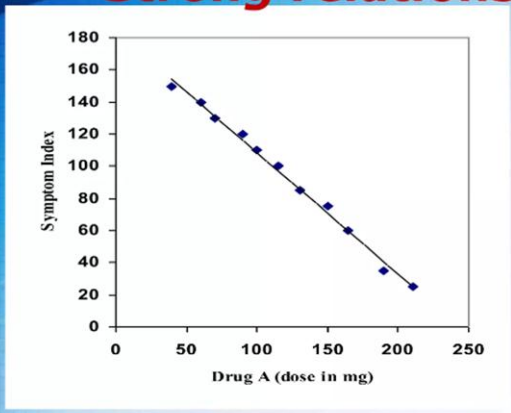
- Two independent and one dependent variable
- One dependent and more than one independent variables
- One dependent variable and more than one independent variable but only one independent variable is considered and other independent variables are considered constant

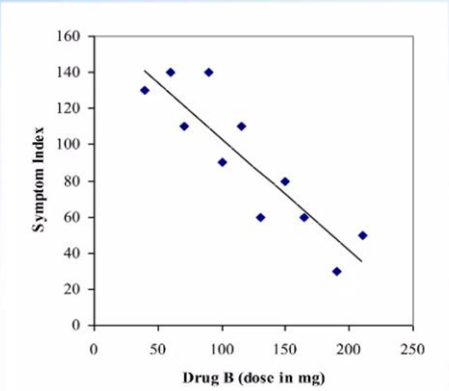
Methods of Studying Correlation

- Scatter Diagram Method
- Karl Pearson Coefficient Correlation of Method
- Spearman's Rank Correlation Method

Correlation: Linear Relationships

Strong relationship = good linear fit





Very good fit

Moderate fit

Points clustered closely around a line show a strong correlation. The line is a good predictor (good fit) with the data. The more spread out the points, the weaker the correlation, and the less good the fit. The line is a REGRESSSION line (Y = bX + a)

Coefficient of Correlation

- **A measure of the strength of the linear relationship between two variables that is defined in terms of the (sample) covariance of the variables divided by their (sample) standard deviations**
- **Represented by "r"**
- **r lies between +1 to -1**
- **Magnitude and Direction**

- $-1 \leq r \leq +1$
- The + and – signs are used for positive linear correlations and negative linear correlations, respectively

Spearmans rank coefficient

- A method to determine correlation when the data is not available in numerical form and as an alternative the method, the method of rank correlation is used. Thus when the values of the two variables are converted to their ranks, and there from the correlation is obtained, the correlations known as rank correlation.

Pearson Correlation Coefficient

- Definition

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}$$

- The value of r always lies between -1 and 1.
- Positive r indicates positive association and negative r indicates negative association.
- The extreme values r = -1 and r = 1 only occur when the data lie exactly on a straight line.
- If the variables have a strong non-linear relationship, r may be small. Always plot the graph.
- High correlation does not prove causality

Correlation and Regression

- Similarities:
  - both dealing with two interval variables
- Differences:
  - Correlation is not causation
  - Regression indicates causal relationship
  - Correlation can't predict
  - Regression can predict

Regression

- Single Linear Regression
  - one independent, one dependent
- Multiple Regression
  - Multiple independent, one dependent
- Logistic Regression

- All dealing with interval variables

Meaning of Index Number

- Index number is a technique of measuring changes in a variable or group of variables with respect to time, geographical location or other characteristics.

Importance

- Index numbers possess much practical importance in measuring changes in the cost of living, production trends, trade, income variations, etc.

Types of Index Numbers

There are many types of index numbers. But there are three basic types in measuring index numbers

- Price Index
- Quantity index
- Value Index

Price Index

- A price index measures the changes in prices from a selected base period to another period

Example

One Kg of Rice is Rs. 20 in the year 1990 (base) and it is Rs. 60 in 2020 (Current Year)

Quantity Index

- A quantity index measures the changes in quantity consumed/Produced from the base period to another period.

Example

A manufacturing industry produces 1000 units of commodity X in the Year 1990 and it produces 5000 units of Commodity X in the Year 2020.

Value Index

- A value index measures the change in the value of one or more items from the base period to the given period.

Value = Price X Quantity

Explanation

In the year 1990, Mr. X Purchase 2 Kg. rice at the rate of Rs. 20 per Kg.

- ==> Value of Rice is Rs. 40 (Rs.20 x 2Kg)

In the year 2020, Mr. X Purchase 2 Kg. rice at the rate of Rs. 60 per Kg.

- ==> Value of Rice is Rs. 120 (Rs.60 x 2Kg)

Note:

The value of the commodity changed from Rs. 40 to 120 ie. base year to current year

THANK YOU

QUESTION BANK

Unit-I&II

1. The word statistics refers to ***Quantitative*** information.
2. Statistics deal with ***Aggregate*** of facts.
3. If numerical facts are to be called statistics, they should be ***Comparable.***
4. Statistics may rightly be called as the ***Science of Averages.***
5. The technique of ***Time Series*** and ***Business Forecasting*** enables the business Men to predict the effect of a large number of variables.
6. The most important limitation of statistics is that it can be ***Misused.***
7. Statistics is a method of ***Decision Making*** in the face of uncertainty based on numerical data and calculated risks.
8. Statistics can be considered both as ***an Art and Science.***
9. ***Comparison*** is one of the main functions of statistics.
10. Statistics helps in making reasonably good ***Forecast.***
11. Statistics help in formulating and testing ***Hypothesis.***
12. One of the important functions ofstatistics is to simplify the ***Complex data*** into diagrammatic and graphic representation.
13. The conclusions obtained statistically are ***not Universally*** true.
14. Statistics are the ***Eyes & Ears***of the state.
15. Statistics is a branch of ***Applied Mathematics*** which specializes in data.
16. Data classified on the basis of some variables that can be measured like this prices, wages, age, number of units produced is called a ***Frequency Distribution.***
17. If a variable is capable on manifesting every conceivable fractional value

within the range of possibilities, it is called ***Continuous*** variable.

18. A ***Discrete*** variable is that which can vary only by finite jumps and cannot manifest every conceivable fractional value.

19. If the lower limit of the first class and the upper limit of the last class are not known, it is called as ***Open – End Class.***

20. Mean, median and mode are ***Same*** in case of symmetrical distribution.

21. The sum of the square of deviation taken from arithmetic mean is ***Minimum.***

22. The sum of deviation taken from arithmetic mean is ***Zero.***

23. The value of ***Median and Mode*** can be determined graphically.

24. If a series is arranged in an ascending or descending order, the median tends to be the value of ***Middle Item.***

25. Median & mode are called ***Positional*** averages.

26. Arithmetic mean is a ***Calculated*** average.

27. The lowest and the highest values that can be included in the class are called ***Class Limits.***

28. The difference between the upper limit and the lower limit of a class is known as ***Class Interval*** of the class.

29. The number of observations corresponding to a particular class is known as the ***Class Frequency.***

30. The midpoint of a class is ***Upper limit + Lower limit / 2.***

31. When the class intervals are so fixed that the upper limit of one class is lower limit of the next class is known as ***Exclusive Method*** of classification.

32. Under ***Inclusive*** method of classification, the upper limit of one class is included in that class itself.

33. To get continuity and to get correct class interval, ***Exclusive*** method of classification is used.

34. Lower limit of 2nd class – upper limit of 1st class / 2 is called as ***Correction Factor.***

35. The number of classes should preferably be between ***5-20.***

36. An attempt to find one single figure to describe the whole of figures is called ***an Average.***

37. The distribution in which two values occur with equal frequency is called ***Bimodal*** distribution.

38. Where mode is ill defined, its value can be calculated by formula ***3 median – 2 mean.***

39. A series which can be described by a discrete variable is called ***Discrete*** series.

40. A series which can be described by a continuous variable is called ***Continuous Series.***

Unit – III

1. The degree of the scatter or variation of the variable about a central value is called ***Dispersion.***
2. Measures of dispersion expressed in the same statistical unit in which the original data are given are called ***Absolute*** dispersion.
3. The ratio of a measure of absolute dispersion to an appropriate average is called the measure of ***Relative*** dispersion.
4. The difference between the value of the smallest item and the value of the largest item included in the distribution is called ***Range.***
5. A high degree of variation would mean ***Little Consistency.***
6. The average difference between the items in a distribution and the median or mean of that series is called ***Mean deviation.***
7. A small standard deviation means ***a high degree of uniformity*** of the observation and homogeneity of series.
8. Algebraic sign are ignored while calculating ***Mean deviation.***
9. The standard deviation is always computed from the ***Arithmetic mean.***
10. For a symmetrical distribution, mean $\pm 1\sigma$ covers ***68.27%*** of the items.

11. The corresponding relative measure of standard deviation is called ***Co – efficient of variation.***
12. The arithmetic mean is more frequently used in calculating the value of ***Average Deviation.***
13. A distribution in which the value of mean, median and mode coincide is called as a ***Symmetric*** distribution.
14. The best measure of central tendency is ***Arithmetic mean.***
15. In case of more than cumulating frequency distribution the frequencies are cumulated ***Upwards.***
16. When the value of mean, median and mode are not equal the distribution is known as ***Asymmetrical or Skewed*** distribution.
17. In a moderately skewed distribution, the distance between the mean and the median is about ***One third*** the distance between the mean and the mode.
18. In case of the cumulative frequency distribution the frequency are cumulated***Downwards.***
19. For studying phenomena like the size of shoes or readymade garment, the appropriate average is ***Mode.***
20. The value which occurs most frequently than others in a series is called ***Mode.***

Unit – IV

1. In case of cumulative series the date are first converted into a ***Simple Series*** either exclusive or inclusive.
2. Studying the relationship between the two variables is called ***Bivariate***analysis.
3. Correlation is said to be negative when the value of two variables move in the ***Opposite Direction.***
4. Correlation is said to be positive when the value of two variables move inthe ***Same Direction.***
5. When we study the relationship between only two variables it is called ***Simple*** correlation.
6. Studying the relationship if only two variables excluding the effect of some other variable is called ***Partial*** correlation.
7. Rank correlation is used to measure correlation between ***Quantitative*** phenomena.
8. If the ratio of change between two variables is uniform, the correlation is said to be ***Linear.***
9. A decrease is one variable is accompanied by decrease in another variable is called ***Positive*** correlation.
10. Studying the relationship between three or more variable simultaneously is called ***Multiple*** correlation.

11. If the ratio of change in one variable does not bear constant ratio of change in another variable, it is called ***Non – Linear*** correlation.
12. Diagrammatic expression of relationship between two variables is known as ***Scatter Diagram.***
13. The value of correlation always lies between ***+1 and -1.***
14. Correlation co-efficient shows the ***Direction*** as well as ***Degree*** of relationship between the variables.
15. The statistical tool that helps to estimate the value of one variable from the given value of another is called ***Regression.***
16. Regression analysis helps as to study the***Nature*** of relationship between variables.
17. The rank correlation co-efficient was developed by ***Spearman.***
18. The variable which is to be predicted is called ***Dependent*** variable.
19. Regression line may also be called as ***Estimating line.***
20. The variable which is the basis for prediction is called ***Independent*** variable.

Unit – V

1. Index numbers are ***Specialized Averages.***
2. The base period should be a ***Normal Period.***
3. ***Fisher's*** index is known as 'ideal'formula for constructing index numbers.
4. Fisher's ideal index satisfies ***Time Reversal Test and Factor Reversal Test.***
5. Index number measures the differences in the magnitude of a group of ***related variables.***
6. The index number for the base period is always taken as ***100.***
7. ***Index numbers*** measure relative changes over successive periods of time.
8. The best average in construction of index number is ***Geometric Mean.***
9. Base year quantities are taken as weight in case of ***Laspeyre'sPrice Index*** method.
10. The specialized average designed to measure the change in group of related variables over a period of time is called ***Index number.***
11. The period against which comparison made is called ***Base*** period.
12. ***Current year*** quantities are taken as a weight in paasche's method.
13. The price of current year expressed as the percentage of the prices of the base year is called ***Price*** relative.
14. The index numbers which measures the average change in the prices paid by the ultimate customers is called***Cost of Living Index.***

15. The index number which studies the net relative change in the volume of goods is called ***Quantity Index.***
16. ***Index numbers*** are one of the most useful devices to know the pulse of the economy.
17. The test to determine whether a given method will work both ways in time, forward and back ward is called ***Time Reversal Test.***
18. The test according to which, the product of a price index and quantity index should be equal to corresponding value index is called ***Factor Reversal Test.***
19. Time reversal test should satisfy the following relation $\mathbf{P_{01} * P_{10} = 1.}$
20. Factor reversal test should satisfy $\mathbf{P_{01} * Q_{01} = p_1 q_1 / p_0 q_0.}$