BUSINESS STATISTICS - II

UNIT-I

REGRESSION ANALYSIS

1) What is regression analysis?

A) Regression Analysis is mathematical measure of the average relationship between two or more variables in terms of the original units of the data. It is used to estimate or predict the unknown value of one variable from the known value of the variable. It is one of the very important statistical tools which is extensively used in almost all sciences- Natural, social and physical sciences. Prediction or estimation is one of the major problems in almost all spheres of human activity. In field of business this tool of statistical Analysis is very widely used. The estimation or prediction of future production, consumption, prices, investments, Sales, Profits, incomes, etc., are of great importance to a businessman.

2) Mention different types of regression?

A) There are different types of regression, some among them are:

i) Simple and Multiple Regression:-

Simple Regression

The Regression Analysis confined to the study of two only two variables at a time is called simple regression.

In such cases, the value of one variable is estimated on the basis of the value of another variable. The variable whose value is to be predicted is called dependent variable and the variable which is used for prediction is called independent variable. In Regression Analysis independent variable is also known as regresser or predictor or explanatory while the dependent variable is also known as regressed or explained variable.

For example if the expenditure on advertisement can have some effect on the volume of sales, then advertising will be independent variable and sales will be dependent variable.

The functional relationship in such cases is expressed as $Y=a+bX$

Multiple Regression

The regression analysis studied on more than two variables at a time is known as multiple regression. A multiple regression analysis on the other hand is one which is made among more than two related variables at a time.

For example, relation between Price, Sales and Income of the people. In such analysis, the value of one variable say, sales are estimated on the basis of other remaining variables i.e., price and income of consumer. In multiple regression one variable is made dependent and other variables are independent.

The functional relationship in such cases is expressed as $Y=f(X, Z)$ or $X= f(Y, Z)$ or $Z= f(X,Y)$

ii) Linear and Non-Linear Regression:-

Linear Regression

A Linear regression analysis is one which gives rise to a straight line where the data relating to the two variables are plotted on a graph paper. This happens, when the two variables have linear relationship with each other which means that with a change in the value of the independent variable by one unit there occurs a constant change in the value of dependent variable. The mathematical equation of such straight line (i.e., Y= a+bX) enables us to study the average change in the value of dependent variable for any given value of the independent variable. A linear trend can be easily projected into the future on the basis of such relationship.

Non-Linear Regression

A non-linear regression analysis, on the other hand is one which gives rise to a curved line when the data relating to two variables are plotted on a graph paper.

3) What are properties of Regression coefficients?

A) The properties of regression coefficients are:

- The correlation coefficient is Geometric Mean between two regression coefficients.
- Both the regression coefficient will have the same sign. The correlation coefficient will have same sign of regression coefficients.
- The arithmetic mean of the two regression coefficient is greater than the correlation coefficient.
- If one of regression coefficient is greater than1, then the other is less than one.

5) Mention the differences between correlation and regression?

A)Correlation	Regression
1. Correlation is a measure of the 'degree and direction' of relationship between the variables.	1. Regression studies 'nature' of relationship between the variables .
2. Correlation does not indicate the cause and effect relationship between the variables.	2. Regression clearly indicates the cause and effect relationship between the variables.
3. Correlation cannot say which variable is the dependent variable and which is independent variable.	3. In regression, the variable corresponding to cause is taken as independent variable and the variable corresponding to effect is taken as dependent variable.
4. Correlation coefficient is a relative measure of the linear relationship.	4. Regression coefficients are absolute e measures indicating the changes in the variable.
5. Correlation analysis cannot be used for predicting or estimating value.	5. Regression analysis is very helpful in predicting and estimating the values.

UNIT-II

INDEX NUMBERS

1) Introduction:- Index Numbers are indicators which reflect relative changes in the level of a certain phenomenon in any given period called the current period with respect to its values in some fixed period called the base period selected for comparison. The variable under consideration may be

(i) The price of a particular commodity like steel, gold, leather, etc., (or) a group of commodities like consumer goods, cereals, milk and milk products cosmetics, etc.,

(ii) Volume of trade, factory production, industrial or agricultural production, imports or exports, stocks and shares, sales and profits of business house and so on.

(iii) The national income of a country, wage structure of worker in various sectors, bank deposits, cost of living of persons of a particular community and so on.

(2) Define Index Numbers and also definitions by different authors?

A) Index Numbers are statistical devices designed to measure the relative change in the level of a phenomenon with respect to time, geographical location or other characteristics such as income, profession, etc., Different experts have defined index numbers in different words. Some of the definitions are

" Index Numbers are devices for measuring differences in the magnitude of a group of related variables"..... CROXTON AND COWDEN

" An Index Numbers is a statistical measure designed to show changes in a variable or a group of related variables with respect to time, geographical location or any characteristic"..... SPEIGES

" An Index Numbers is a statistical device for indicating the relative movement of data where measurement of actual movement is difficult or incapable of being made"..... WELDON

"Index Numbers shows by its variations the changes in magnitude which is not susceptible either of accurate measurement in itself or of direct valuation in practice"..... EDGEWORTH

"Index Numbers are used to measure the change in some quantity which cannot be observed directly, which we know to have a definite influence on many other quantities which can so observe, tending to increase all or diminish all, while this influence is concealed by the action of many causes affecting the separate quantities in various ways"..... BOWLEY

3) Give the uses of Index Numbers?

A) Index Numbers are extensively used for a variety of purposes in Economics, Business, Management, etc., for quantitative data relating to production, consumption, profits, personnel and financial matters as well as for comparing changes in the level of phenomenon for periods, places, etc.,

(i) **Index Numbers are ECONOMIC BAROMETERS**:- Index Numbers are today one of the most widely used statistical devices…They are used as indicators of inflationary or deflationary tendencies".

Like barometers which are used in physics and chemistry to measure atmospheric pressure, Index Numbers are rightly termed as 'economic barometers' or 'barometers of economic activity' which measure the pressure of economic and business behaviour.

(ii) **Index Numbers helps in studying trends and Tendencies**:- Index Numbers study the relative changes in the level of a phenomenon over different periods of time. They are especially useful for study of general trend for a group phenomenon in a time series data.

For example, If a businessman is interested in establishing a new industry, the study of trend, changes in the prices, wages and incomes in different industries is extremely helpful to him to frame a general idea of the comparative courses which the future holds for different undertakings.

(iii) **Index Numbers helps in formulating Decisions and Policies**:- Index Numbers are indispensable for any organization in efficient planning and formulation of executive decisions.

For example, the cost of living Index Numbers are used by the government industrial and business concern for regulation of dearness allowance(D.A) or grant of bonus to workers so as to enable them to meet the increased cost of living from time to time. Similarly Index Numbers are also applied by sociologists, psychologists, health and educational authorities, etc., for formulating and revising their policies from time to time.

(iv) **Index Numbers helps in measuring the purchasing power of money**:- Real wages helps us in determining the purchasing power of money. These are obtained on dividing the money wages by the corresponding price index and multiplying by 100. Cost of living Index Numbers are used to determine whether the real wages are raising or falling.

(v) **Index Numbers are used for Deflation**:- Index Numbers can be deflated to find out the real picture pertaining to the phenomenon. Cost of living Index Numbers are used for deflation of net national product, income value series in national accounts.

For example, obtaining real wages from the given nominal wages, real sales from nominal sales and so on by taking into account appropriate Index Numbers.

(vi) **Index Numbers are used for forecasting**:- Index Numbers provide valuable information that aids forecasting.

For example, an index of sales, along with related indices such as cost of living index, helps in forecasting future demand and future sales of business.

3) Mention the limitations of Index Numbers?

A) Although Index Numbers are very important tools for studying the economic and business activity of a country, they have their limitations and as such should be used and interpreted with caution.

(i) Approximate Representation:- Index Numbers are based on sample data . Hence, they are only approximate indicators. They may not reflect the changes in the relative level of a phenomenon.

(ii) Likelihood of Error:- The likelihood error being introduced at each stage of the construction of the Index Numbers, i.e.,

- Selection of commodities
- Selection of base period
- Collection of the data relating to prices and quantities of the commodities.
- Choice of the formula
- The average to be used for obtaining the index for the composite group of commodities.

(iii) Not Responsive:- There are rapid changes in Technology, tastes, fashions, customs and consequently, the consumption pattern of various commodities. Index Numbers may not be able to keep pace with such changes. Hence they may not be able to reflect the changes in the phenomenon being studied.

4) What are the types of index numbers?

A) Index numbers are indicators. It is constructed for study of any phenomenon. In relation to data pertaining to business and economy, index numbers may be classified into following types:

(i) PRICE INDEX NUMBER:- They measure the changes in prices. They are most common indices. There are various methods by which index numbers can be constructed.

(ii) QUANTITY INDEX NUMBERS :- They study the changes in volume goods produced, consumed or transacted. The index of Industrial production is an example of a quantity index.

(iii) VALUE INDEX NUMBER:- They are rarely used. They study the change in total value, rather than mere changes in prices , change in quantities. They have to be supplemented with price and quantity indices.

5) Explain the methods of constructing index numbers?

A) There are various methods by which index numbers can be constructed. The various methods can be broadly classified into

(i) Unweighted Index Numbers:- In case of unweighted index numbers, all items are given equal importance and equal weights are assigned to all the commodities, i.e., index numbers are calculated basing on prices of commodities.

(ii) Weighted Index Numbers:- In case of weighted index numbers, the index numbers are calculated basing on price and weight of commodities.

UNIT-III

ANALYSIS OF TIME SERIES

1) Define Time Series?

A) Time series is any group of statistical information accumulated at regular intervals of time and arranged in chronological order.

Definition of time series by different scientists:-

A Time Series consists of data arranged chronologically…..*"CROXTON AND COWDEN"*

Time Series may be defined as a collection of readings belonging to different time periods of some economic variable or composite of variables……"YAN-LUN CHOU"

A Time Series is a set of observations taken at specified times, usually at regular intervals……"YAN-LUN CHOU"

Thus Time series is a bi-variate distribution one of the variable is time, the other variable is any phenomenon that one may wish to study. 'Time is an independent variable, while the phenomenon is a function of time i.e., Y = f(t). Time can be expressed in any unit such as year, month, hour, etc., For example, Sales of a company in different years:

Year	1980	1981	1982	1983	1984	1985
Sales	12	16	15	19	21	24

2) What are components of time series? Explain?

A) If the values of a phenomenon are observed at different periods of time, the values so obtained will show appreciable variations or changes. These fluctuations are due to the fact that the value of the phenomenon is affected not by a single factor but due to the cumulative effect of a multiplicity of factors pulling it up and down. The various forces affecting the values of phenomenon in a time series may be broadly classified into the following four categories, known as the components of a time series.

(a) ***Secular Trend*** :- The general tendency of the time series data is to increase or decrease or stagnate during a long period of time called trend. " Trend", also called secular or long-term trend, is the basic tendency of a series to grow or decline over a period of time. The concept of trend does not include short-range oscillations, but rather the steady movement over a long period of time.

An upward trend is usually observed in time series relating to population, production and sales of products, prices, income, money in circulation, etc.,

A downward tendency is noticed in the time series relating to deaths, epidemics, etc., due to advancement in medical technology.

The trend may be linear and Non-linear: If the time series values plotted on the graph cluster more or less round a straight line, the trend exhibited by the time series is termed as Linear trend.

When data of a time series plotted on a graph does not result a straight line, it is said to be non-linear trend.

USES OF TREND :-

i) The long term trend helps us in determining the direction of change.

ii) It helps in forecasting and planning for the future.

iii) It helps in further study of other components and understanding their impact on the phenomenon being studied.
iv) Trend values can be used for extrapolation and where data for certain time periods is missing, it can be used for interpolation.
v) Trend Analysis helps in comparison of two or more time series over different periods of time.

(b) ***Seasonal Variations***:- These variations in a time series are due to the rhythmic forces which operate in a regular and periodic manner over a span of less than a year. The variations involve pattern of change with in a year that tend to get repeated year after year. It can be detected only if the data is recorded in smaller units of time such as weekly, monthly or quarterly, etc., The seasonal variations may be attributed to the following two cases:

(i) Natural Forces :- The various seasons or weather conditions and climatic changes play an important role in seasonal movements. For example, the sales of umbrella pick up very fast in rainy season. The prices of agricultural commodities go down at the time of harvest and then pick up gradually.

(ii) Man-Made conventions :- These variations in a time series with in period of 12 months are due to habits, fashions, customs and conventions of the people in the society. For example, the sales of jewellery and ornaments go up in marriages; the sales and profits in departmental stores go up considerably during festivals.

UTILITY OF SEASONAL VARIATIONS :-

The study and measurement of seasonal patterns constitute a very important part of the analysis of a Time series.
i) Study of seasonal variations help in formulating policy decisions.
ii) Seasonal variations help in planning day to day operations better.
iii) It is helpful in planning diversification strategy.
iv) Seasonal variations helps us to project past pattern into the future.

(*C)* ***Cyclical Variations*** :- The oscillatory movement in a time series with period of oscillations greater than one year are termed as cyclical variations. One complete period which normally lasts from 7 to 9 years is termed as a 'cycle'. 'Business Cycle' has four phases boom(prosperity), recission, depression and recovery from time to time. Most of the economic and business series relating to production, prices, wages, investments, etc., are affected by cyclical upswings and downswings.

The study of cyclical variations is of great importance to business exectutives in the formulation of policies aimed at stabilizing the level of business activity. A knowledge of the cyclic component enables a businessman to have an idea about the periodicity of the booms and depressions and accordingly he can take timely steps for maintaining stable market for his product.

UTILITY OF CYCLICAL VARIATIONS :-
i) Cyclical variations help users to understand that they will not be perennially in the same state of boom or depression.

ii) Cyclical variations is helpful in formulation of policies such as diversification aimed at stabilizing business fluctuations.

(d) ***Random or Irregular Variations*** :- The value of a variable may be completely unpredictable changing in a random manner. These fluctuations are the result of such unforeseen and unpredictable forces that operate in absolutely erratic and irregular manner. They are normally short-term variations caused by non-recurring factors like floods, famines, wars, earthquakes, strikes and lockouts, etc., Irregular variations are also known as episodic fluctuations and include all types of variations in a time series data which are not accounted for by trend, seasonal and cyclical variations.

3) Explain different methods of Measuring Trend?

A) Trend is the basic tendency of a time series to grow or decline over a long period of time. It does not include short term changes and is concerned with steady movement, usually in a particular direction. The various methods that can be used for determining trend are:

(a) FREEHAND OR GRAPHIC METHOD :- In this method, all the values of the given time series are plotted on a graph paper. Then a freehand smoothed curve is drawn joining all the points plotted on the graph. Represent a straight line in such a way that the points of the curve are equally distributed on both the sides of the straight line. This straight line will be the trend line of the given time series data. The trend line drawn by one person is likely to be different from another line drawn by another person for the same data.

Merits:- i) It is very simple.

ii) It does not involve any calculations.

iii) It is very flexible and can be used irrespective of whether the trend is linear or curve-linear.

iv) If used by experienced statisticians, it is a better tool to study trend movement compared to other methods using rigid mathematical formulae.

Demerits:- i) It is very subjective. Different persons may draw different lines and reach different conclusions from the same data. Hence it is not a good forecasting tool.

ii) It requires high levels of experience and expertise to effectively use this method.

(b) METHOD OF SEMI-AVERAGES :- In this method also, all the values of the given time series data are plotted on a graph paper and a free hand smooth curve is drawn joining all the points plotted on the paper.

To represent a trend line, the given data is divided into two parts. If the data is having even number of years, it can be easily divided into two parts. If the data is given for odd number of years, then the year falling exactly in the middle is omitted & for each part an average value is calculated. The two averages so obtained are plotted on the graph against the time period which falls exactly in the middle of each part. The two points are then joined by a straight line which is called trend line. It is extended on either side to forecast the value of the trend for a given time period.

Merits:- i) It is objective. It does not depend on personal judgement.
ii) It is easy to understand and apply.
iii) It can be extended in either direction to obtain earlier estimates and future prediction.

Demerits:- i) It assumes that the trend is linear, which may not be true.
ii) It uses the concept of Arithmetic Mean, which itself has a number of limitations, particularly with regard to influence on extreme values.

(c) METHOD OF MOVING AVERAGES :- This method is very simple and flexible method of measuring trend. It consists in obtaining a series of moving averages of successive overlapping groups or sections of time series data. The Moving Averages is characterized by a constant known as 'period' denoted by 'm'. The calculations of Moving Averages are dependent on the type of period.

Case (i):- When Period is odd--- If m is odd, the successive average of period m is placed opposite to the middle value of the corresponding time interval.

For example, if m is 5, then the average of first successive values are taken and is placed opposite to the third time interval, then leaving the first, add the next 5 values [from 2 to 6] and place the average opposite to 4^{th} time interval & so on. The average so obtained are called trend values.

Case(ii):- When period is even--- If m is even, then there are two middle periods and the successive average is placed in the middle of those two time periods. But the average value will not be opposite to any of the time interval. So, we calculate 2 period moving average where the value will be placed opposite to some time interval. This process is called centering.

For example, If m is 4, then the average of first 4 values are taken and is placed in the middle of 2^{nd} & 3^{rd} time interval leaving the first value, the average of next 4 values[from 2 to 5] is taken and is placed opposite to 3^{rd} & 4^{th} time interval and so on. Again these moving averages will be centered to get a trend value opposite to some time interval.

Note:- If period 'm' is not defined then construct a graph of given time series data. Locate the peak points [the points whose before and after points are small it is called as peak point]. Count the points between each peak points. Take the average of these points that will be the period.

Merits:- (i) It is easy to understand and easy to adopt.
(ii) It is flexible as more observations can be added to given data without affecting trend values already obtained.
(iii) This method is not only used for the measurement of trend but also for measuring seasonal variations, cyclical variations and irregular variations.

De-Merits:- (i) It cannot be used for forecasting or predicting the future trend.
(ii) Trend values cannot be calculated for all the years.
(iii) Moving average values are also greatly affected by extreme values.

(d) METHOD OF LEAST SQUARES:- The method of least squares is the most widely used, most objective and perhaps the best method of obtaining the trend values. In this method the trend equation is obtained by straight line or by parabolic or by exponential curve methods. The equation is fitted using principles of least

squares which minimizes the sum of squares of deviations of actual values from their estimated values. After obtaining the equations of best fit, using that equations we calculate the trend values.

Merits:- (i) It is completely objective with no scope for any personal bias.
(ii) It provides trend values for the entire time period.
(iii) It is useful in forecasting future values.

De-Merits:- (i) It involves many calculations and hence it is tedious.
(ii) The entire calculations need to be re-worked if one single data point is added to a given series.

4) Explain Decomposition of time series data?

A) The analysis of the time series by the process of segregation of its four components. For making proper decomposition of a time series we have three different types of mathematical models before us : They are

(i) **ADDITIVE MODEL**:- In this model it is assumed that the four components of a time series are independent of each other, that none has any effect on the remaining three components, that the observed value is the sum of the four components and that the behavior of the components is of an additive character.

$$Y= T + S + C + I$$

Where Y- observed values T- trend, S – Seasonal variations, c – cyclical variations, I- Irregular Variation.

Under this model, only the absolute values of the other components are added to or subtracted from the trend value to arrive at the observed value of the time series. In practice, the additive model occurs rarely with a time series.

(ii) **MULTIPLICATIVE MODEL**:- In this model, it is assumed that the four components are not necessarily independent and they can affect one another that the behavior of the components is of multiplicative characters and that the geometric mean of all S, C and I would be unity. The observed value of this model is given by $Y = T * S * C * I$

Here S, C, I are not the absolute amounts, but relative variations, are expressed as rates, percentages, or indices fluctuating above or below unity.

The above multiplicative model can also be converted into additive model with the help of logarithms as under:

$$\text{Log } Y = \log T + \log S + \log C + \log I$$

It may be noted that most of the time series relating to business and economics adhere to the multiplicative model in as much as the effect of various factors effecting such time series are not independent of each other.

(iii) **MIXED MODEL**:- Under this model, an observed value of a time series is obtained by any of the following formulae based on the combination of both the additive and multiplicative models:

(i) Y = TSC + I (ii) Y= TS + CI (iii) Y= T + SCI (iv) Y= T + S + CI (v) Y= TC + SI

UNIT-IV

PROBABILITY

In a world we have some things which are certain. For example, we know that every person who is born is going to die. However, there are many things about which we cannot say anything with certainity. For

example if the Indian cricket team is schedule to play a match with a Kenyan team, we cannot say certainly that Indian team will win, Kenyan team also can win. The uncertainty of any thing leads to the study of probability theory.

Experiment:- An operation that have multiple outcomes is called experiment. An experiment is also called trail. For example, tossing of a coin, Rolling a die, etc.,

Event:- The possible outcomes of an experiment is called event.

For example, tossing of a coin is a random experiment and getting head or tail are two separate events.

Exhaustive events:- The total number of possible outcomes in any trail is known as its exhaustive events.

For example, when a die is rolled there are 6 exhaustive events 1,2,3,4,5,6.

Mutually exclusive events:- If the happening of any one of the events. In other words two events are said to be mutually exclusive if both cannot happen simultaneously in a single trail.

For example, if a card is drawn from a pack of 52 cards the event of getting a spade and the event of getting diamond are mutually exclusive.

Equally Likely events:- Events are said to be equally likely when there is no reason to expect any one of them, rather than any one of the others. There is no reason to believe that a particular event has a greater chance of occurring in comparison to another defined event.

For example, when a die is rolled there are six equally likely events.

Independent events:- Two events are said to be independent of each other, if the occurrence of one event does not have any impact on the chance of occurrence of second event.

For example, drawing a card from a pack of cards. Let a king card be drawn from a pack and is not replaced. Another king card is drawn then the second card drawn will have an impact based on first card drawn, then the events are dependent events. If the card is replaced before drawing the second card then there will be no impact on the second card drawn, then the events are said to be independent events.

Simple events or Elementary events:- The events which cannot be split up further into sub events is called simple event. In other words, the event which have only one outcome are called simple events or elementary events.

For example, tossing of a coin, the outcome will be either head or tail.

Compound events:- The event which can be further divided into sub-events is called compound event.

For example, getting an even number on the die, when a die is rolled we can get an even or an odd number. The even number event can be sub-divided into 2, 4, 6.

Sample Space:- The set of all possible outcomes of an experiment is called sample space. It is denoted by S.

For example, when two coins are tossed there will be four outcomes, then the sample space is

S={HH, HT, TH, TT}

ADDITION THEOREM OF PROBABILITY:-

If A & B are two events, then the probability of occurrence of atleast one event is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If A & B are two mutually exclusive events, then

$$P(A \cup B) = P(A) + P(B)$$

If A , B & C are three events, then the probability of occurrence of atleast one event is

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$$

$P(A^C) = 1 - P(A)$

$P(A^C \cup B^C) = P(A^C) + P(B^C) + P(A^C \cap B^C)$

$P(A^C \cap B) = P(B) - P(A \cap B)$

$P(A \cap B^C) = P(A) - P(A \cap B)$

Demorgan laws:- (i) $P(A^C \cup B^C) = P(A \cap B)^C$ (ii) $P(A^C \cap B^C) = P(A \cup B)^C$

MULTIPLICATION THEOREM OF PROBABILITY:-

Compound Probability:- The simultaneous occurrence of events is called compound probability.

If A & B are two events then $P(A \cap B)$ is compound probability.

Conditional Probability:- The occurrence of an event when the other event has already occurred is called conditional Probability.

If A & B are two events then the occurrence of event A when B has already occurred is defined as P(A/B). The occurrence of event B when A has already occurred is defined as P(B/A).

Multiplication Theorem:-

If A & B are two events then the simultaneous occurrence of the events is the product of unconditional probability of event A and conditional probability of event B and Vice versa. i.e.,

$P(A \cap B) = P(A) * P(B/A)$

$P(A \cap B) = P(B) * P(A/B)$

If A & B are two independent events then $P(A \cap B) = P(A) * P(B)$

Bayees Theorem:

Let E1,E2,……,En be a set of events associated with a sample space S, where all the events E1,E2,……,En have non zero probability of occurrence and they form a partition of S. Let A be any event associated with S, then according to Bayes' theorem,

$$P[E_i / A] = \frac{P[E_i].P[E_i / A]}{\sum P[E_i].P[E_i / A]}$$

UNIT-V

THEORETICAL DISTRIBUTIONS