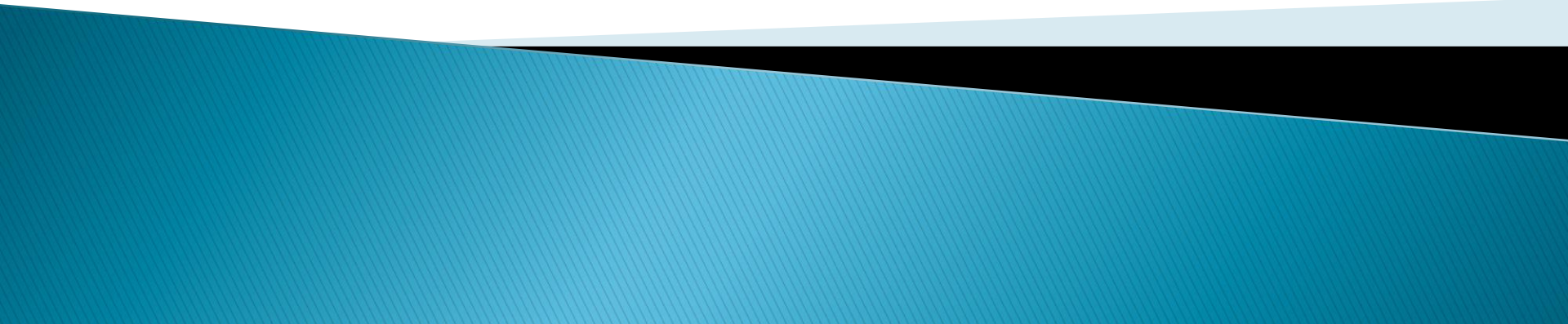


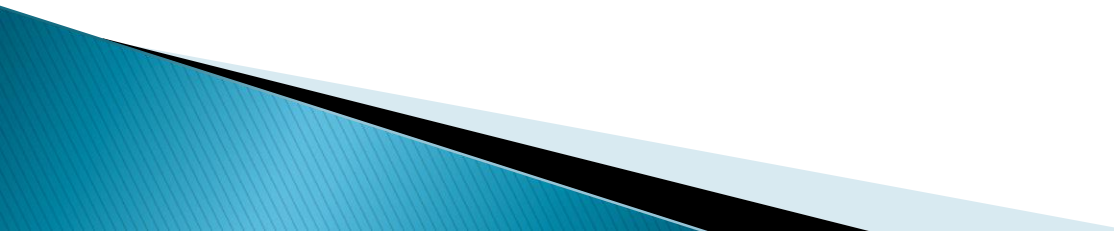
BIOSTATISTICS

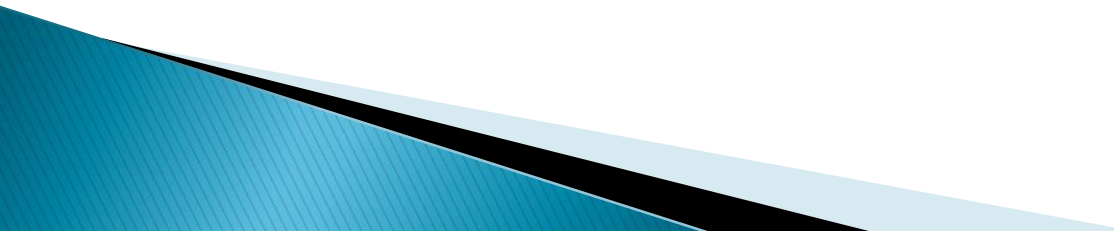
CORRELATION AND REGRESSION

Gincy C George
(Assistant Professor On Contract)

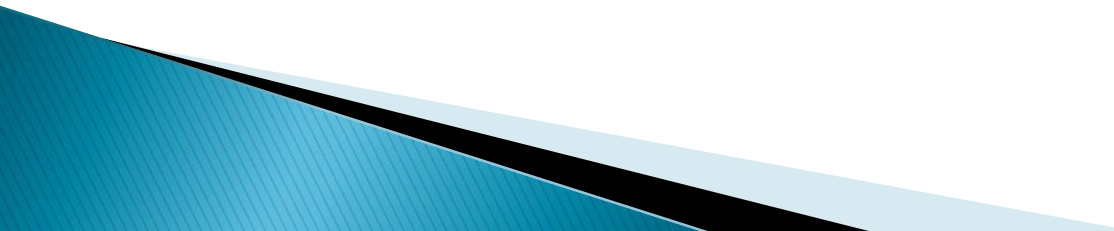
CORRELATION AND REGRESSION



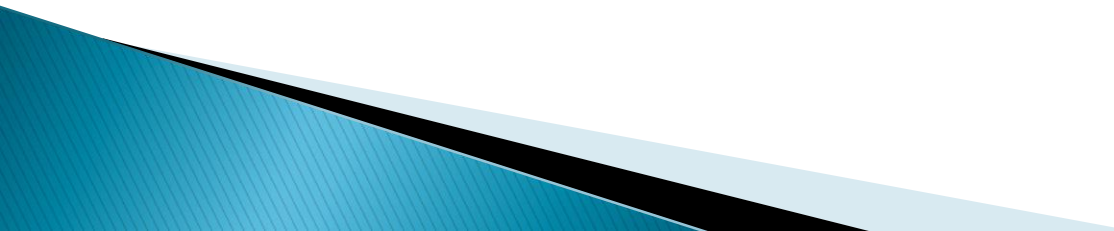
- ▶ Correlation is basically a measure of relationship between two variables.
 - ▶ Two variables are said to be correlated if an increase in one variable is an average accompanied by an increase (or decrease) of the other.
 - ▶ In another words, correlation is said to exist when the two groups or series of items vary together either directly or inversly.
- 

- ▶ The measure of correlation is called correlation coefficient or correlation index.
 - ▶ It summarizes in one figure the direction and degree of correlation.
 - ▶ The correlation analysis refers to the statistical techniques used in measuring the closeness of the relationship between the variables.
- 

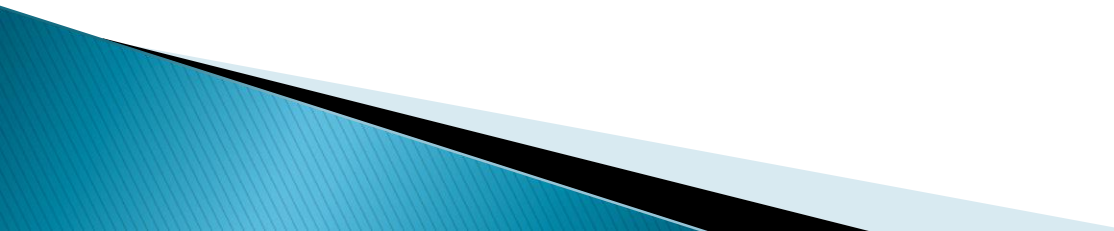
TYPES OF CORRELATION

- ▶ When the movement of the two variables are in the same direction, the correlation said to be **positive or directive**.
 - ▶ When the movement of the two variables in the opposite direction the correlation said to be **negative or inverse**
- 

Simple and multiple correlation

- ▶ When only two variables are studied, it is a case of **simple correlation**. Eg; correlation between the height and weight of the individuals.
 - ▶ In **multiple correlation**, there are three or more variables are studied simultaneously. Eg; yield of rice, hector, amount of rainfall and the amount of fertilizers used.
- 

METHODS TO MEASURE CORRELATION

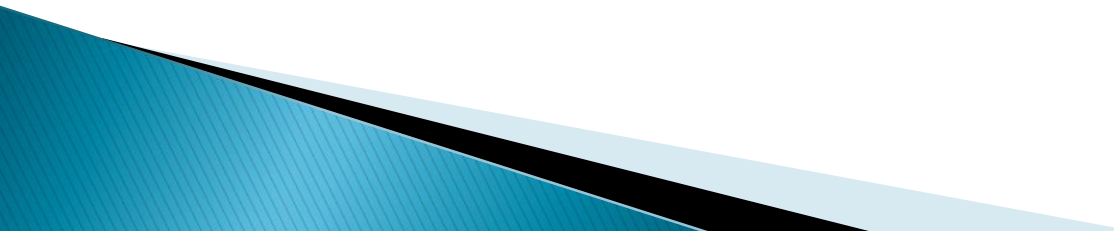
- ▶ 1. SCATTER DIAGRAM METHOD
 - ▶ 2.KARL PEARSON'S COEFFICIENT OF CORRELATION
 - ▶ 3.SPEARMAN'S CORRELATION
- 

SCATTER DIAGRAM METHOD

- ▶ This is the simplest method of finding out whether there is any relationship present between two variables.
- ▶ It is studied by plotting the values on a graph paper.
- ▶ And the graph is known as scatter diagram

▶ Eg;

x	Y
5	3
10	5
15	6
20	9
25	10

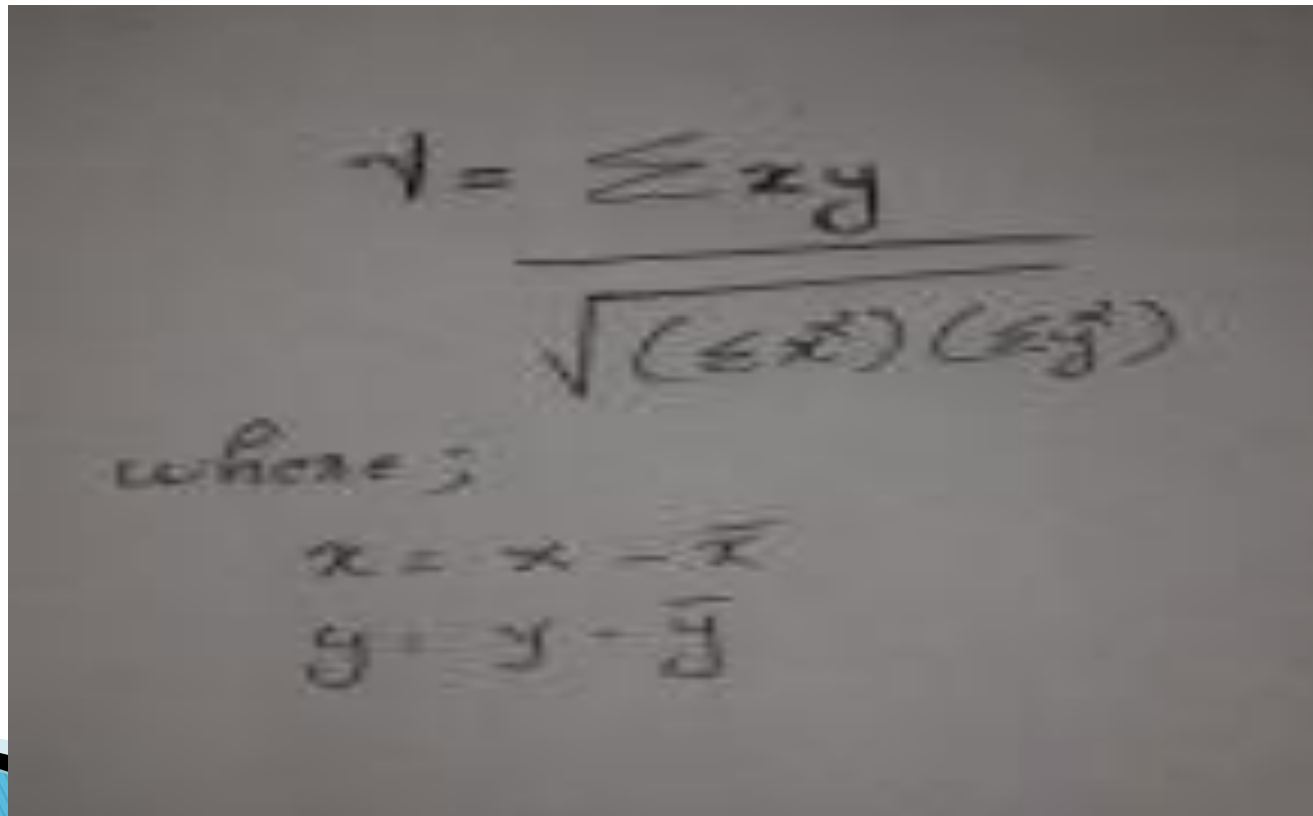
- ▶ If the plotted points form a straight line running from the lower left hand corner, then there is perfect positive correlation.
 - ▶ On the other hand, if the points are in a straight line having a falling trend from the upper left hand corner to the lower right hand corner, it reveals that there is a perfect negative correlation.
- 

KARLPEARSON'S COEFFICIENT OF CORRELATION

- ▶ Of the several statistical methods of measuring correlation the karlpearson's method, popularly known as the pearson's coefficient of correlation is the most widely used in practice.
- ▶ It is denoted by the symbol ' r '

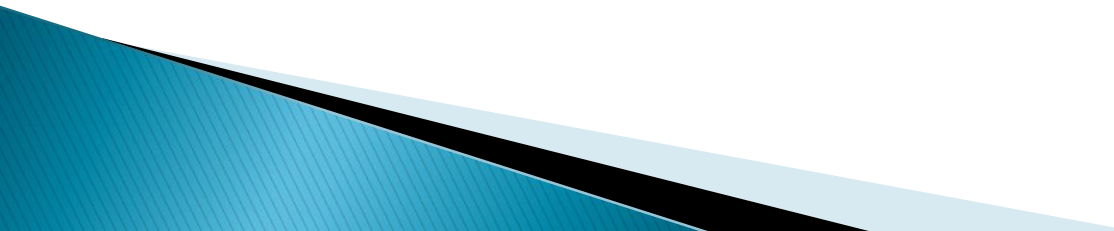
calculation

- ▶ The formula for computing pearson's 'r' is;



A photograph of a piece of paper with handwritten mathematical formulas. The top formula is the Pearson correlation coefficient:
$$r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}$$
 Below this, the word "where;" is written. Then, the deviation scores for x and y are defined:
$$x = x - \bar{x}$$

$$y = y - \bar{y}$$

- ▶ The value of 'r', as obtained by the above formula shall always be between ± 1 .
 - ▶ Where; $r = +1$, it means perfect positive correlation between the variables.
 - ▶ And if $r = -1$, it means perfect negative correlation between the variables,
 - ▶ When $r = 0$, it means there is no relationship between the two variables.
- 

$$\bar{E}^2 : \bar{E}^2 = \frac{(\sum E)^2}{N}$$

$$\bar{y}^2 : \bar{y}^2 = \frac{(\sum y)^2}{N}$$

$$\bar{xy} : \bar{xy} = \frac{(\sum x)(\sum y)}{N}$$

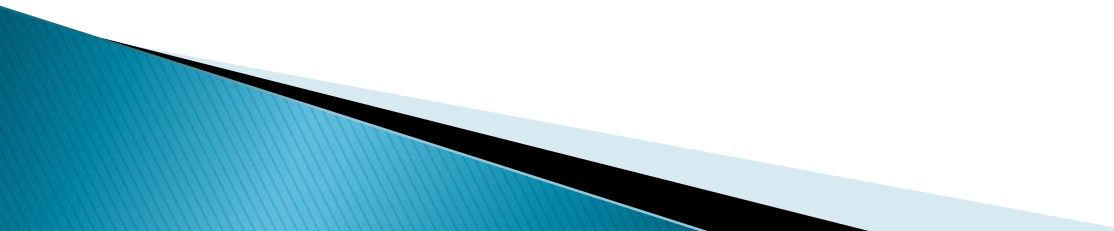
problem

- ▶ The following table gives the age(years) of husbands and their wives. Calculate the pearson's coefficient of correlation and comment on your result.

Sl.no of pairs	1	2	3	4	5	6	7	8	9	10
Age of husbands	25	27	29	32	35	39	42	44	44	50
Age of wives	20	21	24	28	30	33	37	40	43	46

sl.no	Age of husbands(X)	Age Of wives(Y)	x2	y2	XY
1	25	20			
2	27	21			
3	29	24			
4	32	28			
5	35	30			
6	39	33			
7	42	37			
8	44	40			
9	44	43			
10	50	46			

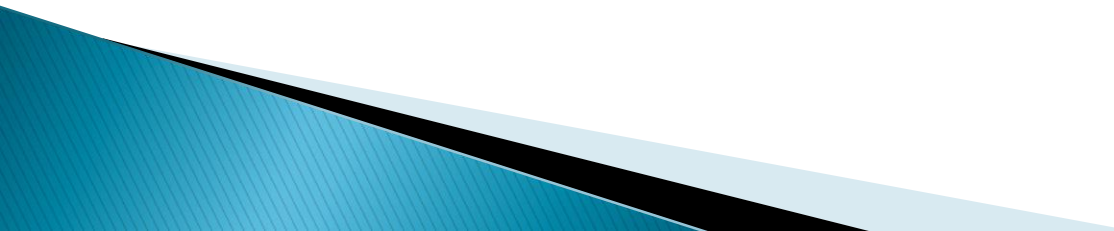
SPEARMAN'S RANK COEFFICIENT OF CORRELATION

- ▶ In 1904, Charles Edward Spearman, an English psychologist and statistician found out a method of ascertaining the coefficient of correlation by ranks.
 - ▶ This method is useful in dealing with qualitative characteristics such as intelligence, beauty, efficiency etc.
 - ▶ Rank correlation is applicable only to individual observations.
- 

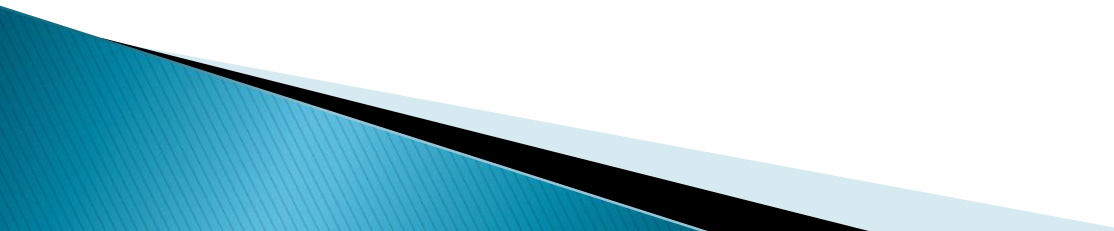
- ▶ The result we get from this method is only an approximate 1. because under ranking method the original value are not taken into account.
- ▶ Spearman coefficient of correlation is designated by the symbol 'ρ' (rho), and is calculated using the following formula;

formula

$$\rho = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$
$$\rho = 1 - \frac{6 \sum D^2}{N^3 - N}$$

- ▶ The value of rho lies between +1 and -1.
 - ▶ If $\rho = +1$ then there is complete agreement in the order of ranks and the direction of the rank is also the same.
 - ▶ When rho is -1 then there is complete disagreement in the order of ranks and they are in opposite directions.
 - ▶ There are two types of problems; one where ranks are given, 2, where ranks are not given.
- 

When ranks are given

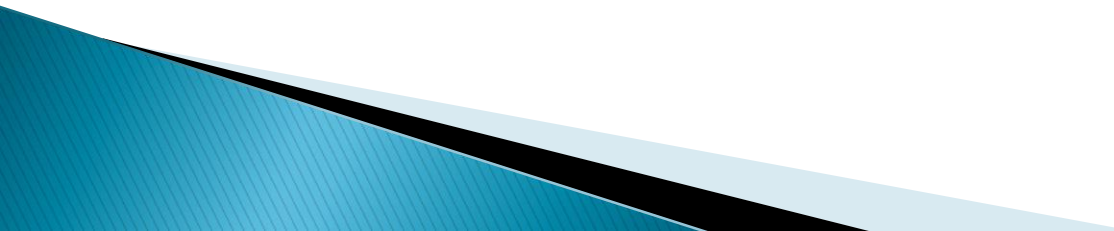
- ▶ Where the actual ranks are given, the steps are as follows;
 - ▶ Compute the difference of the two ranks (R_1 & R_2) and denote it by D .
 - ▶ Square D
 - ▶ Substitute in the given formula.
- 

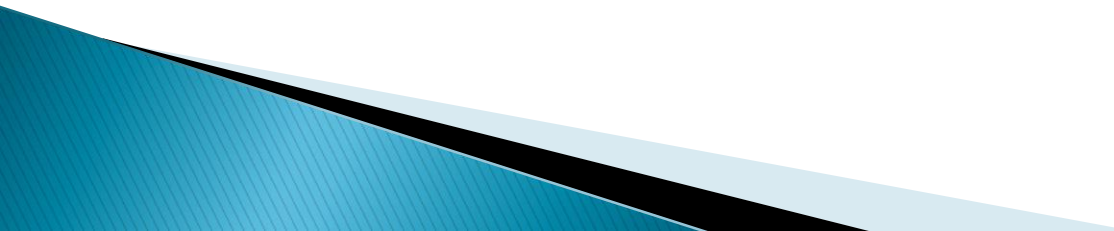
problem

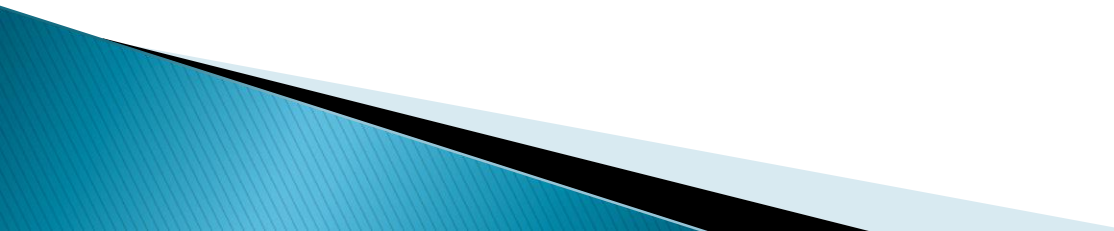
- ▶ The following are the ranks obtained by ten students in two subjects, biostatistics and biochemistry. Compute the spearman rank coefficient of correlation and discuss your result.

Sl.no	1	2	3	4	5	6	7	8	9	10
Biostatistics	7	2	1	10	8	4	9	6	3	5
biochem	9	1	2	10	7	6	5	8	4	3

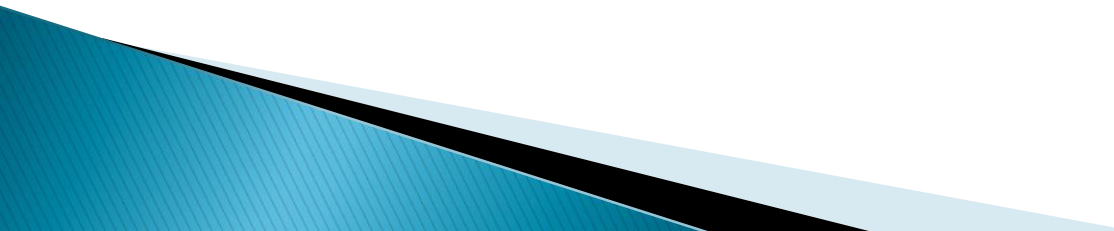
REGRESSION ANALYSIS

- ▶ A central problem in statistical inference involving a bivariate distribution is to determine the true relationship between X and Y . ie, how the two variables behave in relation to each other.
 - ▶ The statistical tool with the help of which are in a position to estimate(predict) the unknown values of one variable from known values of another is called regression analysis.
- 

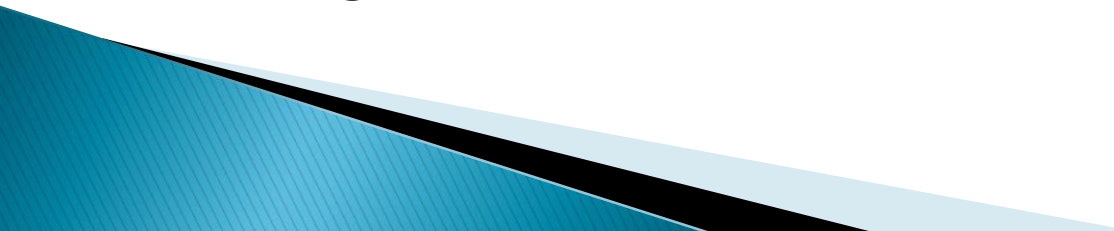
- ▶ Regression reveals an average relationship between two variables and thus makes possible estimation or prediction.
 - ▶ The prediction is made possible because a cause and effect relation is assumed to exist between two or more variables.
 - ▶ The variable that is the basis of estimation is conveniently called the independent variable and is designated X , and the variable whose value is to be estimated is called the dependent variable and is designated Y .
- 

- ▶ When an equation is formulated for estimating Y from X, such as an equation is called regression is called regression of Y on X.
 - ▶ Regression analysis is generally classified into two types. 1.simple and 2.multiple.
 - ▶ Simple regression involves 2 variables, such as regression of Y on X.
 - ▶ Multiple regression involves 3 or more variables. One of which is the dependent variable that is to be associated with the values of all other.
- 

Differences between correlation and regression analysis

- ▶ 1. correlation coefficient is a measure of covariability between X and Y, the objectives of regression analysis is to study the nature of relationship between the variable so that we may be able to predict the value of one on the basis of the other.
 - ▶ 2. the cause and effect relation is clearly indicated through regression analysis than in correlation.
- 

Linear regression

- ▶ Variables have either linear or non linear relationship.
 - ▶ Two variables are said to have a linear relationship when change in the independent variable by one unit leads to a constant absolute change in the dependent variable.
 - ▶ When 2 variables have linear relationship, the regression lines can be used to find out the values of dependent variable.
 - ▶ Ie, the equation describing the relation between X and Y is linear and is graphically represented by a straight line
- 

REGRESSION EQUATION

- ▶ The equation that describes position of any line on a graph is called regression equation. For a linear regression, the equation for a dependent variable Y against independent variable X can be given as follows;
- ▶ $Y = a + bx$
- ▶ Here , values of “a” and “b” are constant and are fixed for a particular line. If the values of ‘a’ and ‘b’ are known, y can be obtained for any corresponding value of x .
- ▶ The ‘a’ and ‘b’ are coefficients.

THANK YOU