INTRODUCTION

Correlation Analysis and Regression Analysis are the two analyses that are made from a multivariate distribution i.e. a distribution of more than one variable. In particular when there are two variables, say x and y, we study bivariate distribution. We restrict our discussion to bivariate distribution only.

Correlation analysis, it may be noted, helps us to find an association or the lack of it between the two variables x and y.

Regression analysis, on the other hand, is concerned with predicting the value of the dependent variable corresponding to a known value of the independent variable on the assumption of a mathematical relationship between the two variables and also an average relationship between them.

BIVARIATE DATA

When data are collected on two variables simultaneously, they are known as bivariate data and the corresponding

CORRELATION AND REGRESSION

frequency distribution, derived from it, is known as Bivariate Frequency Distribution.

As in the case of a Univariate Distribution, we need to construct the frequency distribution for bivariate data. Such a distribution takes into account the classification in respect of both the variables simultaneously.

- "Marginal Distribution". It is the frequency distribution of one variable (x or y) across the other variable's full range of values.
- "Conditional Distribution". It is the frequency distribution of one variable (x or y) across a particular sub-population of the other variable.

CORRELATION ANALYSIS

the change in one variable is reciprocated by a corresponding change in the other variable either directly or inversely, then the two variables are known to be associated or correlated. Otherwise, the two variables are known to be dissociated or uncorrelated or independent.

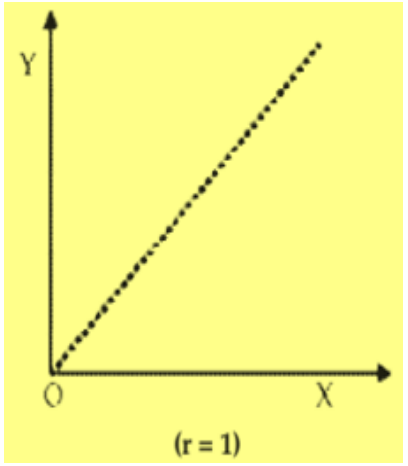
There are two types of correlation.

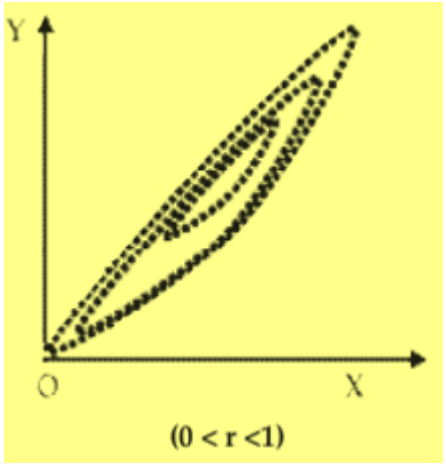
- Positive correlation
- Negative correlation

MEASURES OF CORRELATION

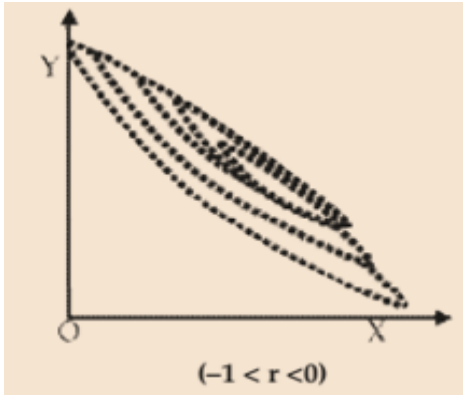
SCATTER DIAGRAM

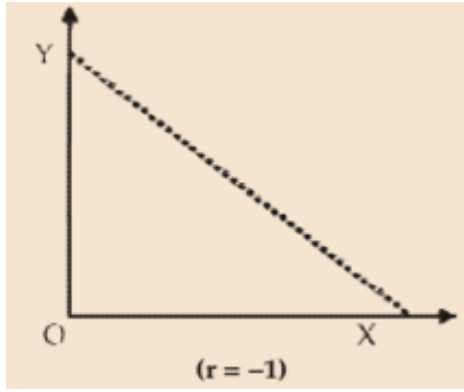
This is a simple diagrammatic method to establish correlation between a pair of variables. Unlike product moment correlation co-efficient, which can measure correlation only when the variables are having a linear relationship, scatter diagram can be applied for any type of correlation – linear as well as non-linear i.e. curvilinear.

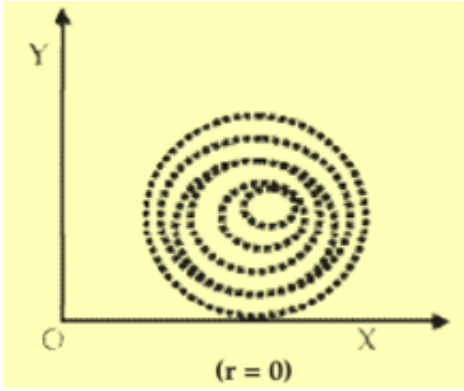


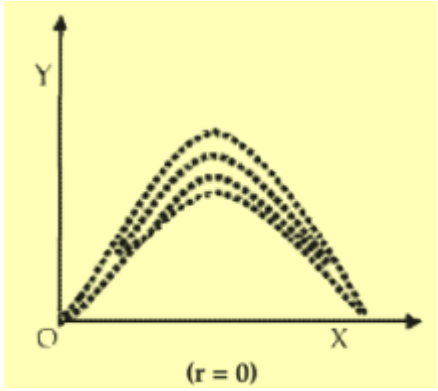


CORRELATION AND REGRESSION









KARL PEARSON'S PRODUCT MOMENT CORRELATION COEFFICIENT

This is by for the best method for finding correlation between two variables provided the relationship between the two variables is linear. Pearson's correlation coefficient may be defined as the ratio of covariance between the two variables to the product of the standard deviations of the two variables.

- $$\frac{n\sum xy-\sum x\sum y}{\sqrt{n\sum x^2-(\sum x^2)}\times\sqrt{n\sum y^2-(\sum y)2}}$$

PROPERTIES OF CORRELATION COEFFICIENT

- The Coefficient of Correlation is a unit-free measure.
- The coefficient of correlation remains in value, not necessary in sign under a change of origin and/or scale of the variables under consideration depending on the sign of scale factors.
- The coefficient of correlation always lies between –1 and 1, including both the limiting values i.e.

CORRELATION AND REGRESSION

SPEARMAN'S RANK CORRELATION COEFFICIENT

Rank correlation can also be applied to find the level of agreement (or disagreement) between two judges so far as assessing a qualitative characteristic is concerned.

$$r_R = 1-\frac{6\sum d^2 i}{(n^2-1)}$$

$$r_R = 1-\frac{6\left[\sum_i d_i^2+\sum_j \frac{\left(tj^3-t_j\right)}{12}\right]}{n\left(n^2-1\right)}$$

COEFFICIENT OF CONCURRENT DEVIATIONS

A very simple and casual method of finding correlation when we are not serious about the magnitude of the two variables is the application of concurrent deviations. This method involves in attaching a positive sign for a x-value (except the first) if this value is more than the previous value and assigning a negative value if this value is less than the previous value.

COEFFICIENT OF CONCURRENT DEVIATIONS

$$r_c = \pm\sqrt{\pm\frac{(2c-m)}{m}}$$

REGRESSION ANALYSIS

Regression analysis plays a very important role in the field of every human activity. A businessman may be keen to know what would be his estimated profit for a given level of investment on the basis of the past records.

PROPERTIES OF REGRESSION LINES

- The regression coefficients remain unchanged due to a shift of origin but change due to a shift of scale.
- The two lines of regression intersect at the point x,y , where x and y are the variables under consideration.
- The coefficient of correlation between two variables x and y in the simple geometric mean of the two regression coefficients. The sign of the correlation coefficient would be the common sign of the two regression coefficients

CORRELATION AND REGRESSION

PROBABLE ERROR

Probable Error is a method of obtaining correlation coefficient of population. It is defined as

$$P.E = 0.674 \times \frac{1-r^2}{\sqrt{N}}$$

$$PE = \frac{2}{3}\ SE$$

$$SE = \frac{1-r^2}{\sqrt{N}}$$

The following are the assumption while probable Errors are significant.

- If r< PE there is no evidence of correlation
- If the value of 'r 'is more than 6 times of the probable error, then the presence of correlation coefficient is certain
- Since 'r 'lies between -1 and +1 (-1 < r < 1) the probable error is never negative.

CORRELATION AND REGRESSION

REVIEW OF CORRELATION AND REGRESSION ANALYSIS

The 'coefficient of non-determination' is given by (1–r2) and can be interpreted as the ratio of unexplained variance to the total variance

Coefficient of non-determination = (1–r2)

The two lines of regression coincide i.e. become identical when r = –1 or 1 or in other words, there is a perfect negative or positive correlation between the two variables under discussion. If r = 0 Regression lines are perpendicular to each other