CORRELATION

"A measure association between two numerical variables is known as correlation".

In other words "Correlation is the relationship that exists between two or more variables. If variable are related in such a way that change in one creates a corresponding change in the other, then the variable are said to be correlated".

Examples Of Correlation:

- ✓ Relationship between height and weight.
- ✓ Relationship between price of a commodity and demand of commodity.
- ✓ Relationship between the advertisement and Sales

Need to study the correlation:

- ✓ Correlation help in study economic theory and business studies, it help in establishing relationship between variable like price and quantity demanded, advertising and sales promotion measures.
- ✓ Correlation analysis helps in deriving precisely the degree and direction of such relation.
- ✓ The effect of correlation is to reduce the range of uncertainty of our prediction. The prediction based on correlation analysis will be more reliable and near to reality.
- ✓ The measure of coefficient of correlation is relative measure of change.

TYPES OF CORRELATION

There are three broad types of correlation.

- ✓ Positive and negative
- ✓ Linear and non – linear
- ✓ Simple, partial and multiple

Positive and Negative Correlation:

- **Positive correlation:** If both the variables vary in same direction, correlation is said to be positive correlation, i.e. if one variable increase, the other also increases or, if one variable decrease, the other variable is also decreases said to be a positive correlation.
  Eg. Relationship between height and weight.

- **Negative correlation:** If both the variable vary in opposite direction, the correlation is said to be negative. In other word if one variable increases, other variable decreases or if one variable decreases , the other variables increases, than correlation between two variables is said to be negative correlation.

Eg. Relationship between price of a commodity and demand of commodity.

Linear and Non – Linear Correlation:

- **Linear Correlation:** If the amount of changes in one variable bears a constant ratio to the amount of changes in the other variable, then correlation is said to be linear. If such variable plotted on graph paper all the plotted points would fall on straight line.

  Eg. Relationship between weight of cheese (in kg) and amount of milk (in liter).

- **Non-Linear Correlation :** If change in one variable does not bear a constant ratio to the amount of changes in the other variable, then correlation is said to be non-linear. If such variable plotted on graph, the point will fall on a curve and not a straight line.

  Eg. Relationship between sales (in Rs.) and advertisement expenses.

Simple, Partial and Multiple Correlations:

- **Simple correlation:** In simple correlation we study only two variables – say price and demand.
- **Multiple correlation:** In multiple correlations we study together the relationship between three or more factors like production, rainfall and use of fertilizers.
- **Partial correlation:**In partial correlation though more than two factors are involved but correlation is studied only between two factors and the other factors are assumed to be constant.

MEASURING THE RELATIONSHIP

Karl Pearson's Correlation Coefficient (r): Measure the direction and the strength of the linear association between two numerical paired variables.

Karl Pearson's coefficient of correlation "r" is given by:

$$r = \frac{\sum xy}{N\, \sigma_x\, \sigma_y}$$

Where $x = (X - \bar{X})$

$y = (Y - \bar{Y})$

$$\sigma_x = Standard\ devitation\ of\ series\ X = \sqrt{var\ X} = \sqrt{\frac{\sum(X - \bar{X})^2}{N}}$$

$$\sigma_y = Standard\ devitation\ of\ series\ Y = \sqrt{var\ Y} = \sqrt{\frac{\sum(Y-\bar{Y})^2}{N}}$$

$$N = Number\ of\ observation$$

$$r = \frac{Covariance\ (X,Y)}{\sqrt{var\ X}\sqrt{var\ Y}}\ where\ Covariance = \frac{\sum xy}{N}$$

$$r = \frac{Cov\ (X,Y)}{\sigma_x\ \sigma_y} = \frac{\sum xy}{\sqrt{\sum x^2}\sqrt{\sum y^2}}$$

Interpretation of Correlation Coefficient (r) :

Value of r	Interpretation
1	Perfect positive correlation
0	No correlation
-1	Perfect negative correlation

Properties of the Linear Correlation Coefficient (r):

- ✓ $-1 \le r \le 1$
- ✓ Value of *r* does not change if all values of either variable are converted to a different scale.
- ✓ The *r* is not affected by the choice of *x* and *y*. Interchange *x* and *y* and the value of *r* will not change.
- ✓ *r* measures strength of a linear relationship.
- ✓ Correlation coefficient gives direction as well as degree of relationship between the variables.
- ✓ Correlation coefficient along with other information help in estimating the values of the dependent variable from the known value of independent variable.

Limitation of Karl Person's coefficient of correlation:

- ✓ Assumption of linear relationship: The assumption of linear relationship between the variables may or may not hold always true .
- ✓ Time consuming:-Its computation is time consuming as compared to other method.
- ✓ Affected by extreme values: -It is affected by the value of extreme item.
- ✓ Requires careful interpretation:-The investigation should reach a conclusion based on logical reasoning and investigation on significantly related matter.

STANDARD ERROR AND PROBABLE ERROR

$$S.E.(r) = \frac{1-r^2}{\sqrt{N}}$$

$$P.E.(r) = 0.6745\ \times S.E. = 06745\ \times \frac{1-r^2}{\sqrt{N}}$$

PARTIAL CORRELATION

- ✓ Coefficient of partial correlation between X_1 and X_2 keeping X_3 constant is given by

$$r_{12.3} = \frac{r_{12} - r_{13}.r_{23}}{\sqrt{1-r_{13}{}^2}\ \sqrt{1-r_{23}{}^2}}$$

- ✓ Coefficient of partial correlation between X_2 and X_3 keeping X_1 constant is given by

$$r_{23.1} = \frac{r_{23} - r_{12}.r_{13}}{\sqrt{1-r_{12}{}^2}\ \sqrt{1-r_{13}{}^2}}$$

- ✓ Coefficient of partial correlation between X_1 and X_3 keeping X_2 constant is given by

$$r_{13.2} = \frac{r_{13} - r_{12}.r_{23}}{\sqrt{1-r_{12}{}^2}\ \sqrt{1-r_{23}{}^2}}$$

MULTIPLE CORRELATION

- ✓ Coefficient of multiple correlation between X_1, X_2 and X_3 treating X_1 as dependent and X_2 and X_3 as independent is given by

$$R_{1.23} = \frac{\sqrt{r_{12}{}^2 + r_{13}{}^2 - 2r_{12}.r_{13}.r_{23}}}{1-r_{23}{}^2}$$

- ✓ Coefficient of multiple correlation between X_1, X_2 and X_3 treating X_2 as dependent and X_1 and X_3 as independent is given by

$$R_{2.13} = \frac{\sqrt{r_{12}{}^2 + r_{23}{}^2 - 2r_{12}.r_{13}.r_{23}}}{1-r_{13}{}^2}$$

- ✓ Coefficient of multiple correlation between X_1, X_2 and X_3 treating X_1 as dependent and X_2 and X_3 as independent is given by

$$R_{3.12} = \frac{\sqrt{r_{13}{}^2 + r_{23}{}^2 - 2r_{12}.r_{13}.r_{23}}}{1-r_{12}{}^2}$$

REGRESSION

REGRESSION

If two variables are significantly correlated, and if there is some theoretical basis for doing so, it is possible to predict values of one variable from the other. This observation leads to a very important concept known as 'Regression Analysis'.

In other words "The statistical technique to find an algebraic relationship between two or more variables in the form of an equation to estimate the value of a random variable, given the value of another variable, is called regression analysis. "

Variable:

Independent and Dependent Variable: There are two types of variables in a regression analysis:

- ✓ The independent variable is the variable in regression that can be controlled or manipulated.
- ✓ The dependent variable is the variable that cannot be controlled or manipulated.

DIFFERENCE BETWEEN CORRELATION AND REGRESSION

- ✓ The process of developing an algebraic equation between two or more variables from sample data and predicting the value of one variable given the value of other variable is referred to as regression analysis, while measuring the strength of the relationship between two or more variables is referred as correlation analysis.
- ✓ Correlation doesn't has a cause – effect relationship, while regression has the cause – effect relationship.
- ✓ Correlation coefficient is independent of change of scale and origin. Regression coefficient is independent of change of origin but note scale.
- ✓ In regression analysis there is only one dependent variable, while in correlation analysis there is no such restriction.

REGRESSION LINES

There are two regression lines which shows the algebraic relationship between two variables. These lines may be in the form of straight line or parabolic curve.

- ✓ Regression line of Y on X and
- ✓ Regression line of X on Y.

Regression line of Y on X: The Regression line of Y on X is expressed as follows:

$$Y = a + bX$$

In this equation Y is treated as dependent variable and X is treated as independent variable.

Here *a* and *b* are numerical constants. '*a*' is "Y - intercept" and '*b*' is slop of the line.

Regression line of X on Y: The Regression line of Y on X is expressed as follows:

$$X = a + bY$$

In this equation X is treated as dependent variable and Y is treated as independent variable.

Here *a* and *b* are numerical constants. '*a*' is "X - intercept" and '*b*' is slop of the line.

Aims Tutorial
Oriental Bank of Commerce
H.No.:69/339 VT Road Ward 27,
Mansarovar Jaipur (Rajasthan) 302020
Contact No: +91-8947920041

METHOD OF FINDING THE REGRESSION LINE

There are two method of finding the equations of regression line.

- ✓ Mean based method
- ✓ Least square method

Mean Based Method: Equations of the two regression lines based on means are represented as follows:

$$\textbf{X on Y}: \quad (X-\bar{X}) = b_{xy}\ (Y-\ \bar{Y})$$

$$\textbf{Y on X}: \quad (Y-\bar{Y}) = b_{yx}\ (X-\ \bar{X})$$

Where $\bar{X} = Mean\ of\ series\ X$

$\bar{Y} = Mean\ of\ series\ Y$

$$b_{xy} = regression\ coefficient\ of\ X\ on\ Y\ = \frac{\sum xy}{\sum y^2}$$

$$b_{yx} = regression\ coefficient\ of\ Y\ on\ X\ = \frac{\sum xy}{\sum x^2}$$

Least Square Method: The values of parameters 'a' and 'b' in the regression lines can be obtained by solving the two simultaneous linear equations, these equations are known as normal equations.

- **Equation of regression lines Y on X:**

  We know that the Regression line of Y on X is expressed as follows:

  $$Y = a + bX$$

  The normal equations for finding the value 'a' and 'b' are:

  $$\sum Y = Na + b\sum X \quad \text{and}$$

  $$\sum X\,Y = a\sum X + b\sum X^2.$$

- **Equation of regression lines X on Y**:

  We know that the Regression line of X on Y is expressed as follows:

  $$X = a + bY$$

  The normal equations for finding the value of 'a' and 'b' are:

  $$\sum X = Na + b\sum Y \quad \text{and}$$

$$\sum XY = a\sum Y + b\sum Y^2.$$

REGRESSION COEFFICIENT

The regression coefficient is the slope of the regression line and tells you

- ✓ What the nature of the relationship between the variables is.
- ✓ How much change in the independent variables is associated with how much change in the dependent variable.
- ✓ The larger the regression coefficient the more change.

There are two regression coefficients:

- ✓ Regression coefficient of X on Y
- ✓ Regression coefficient of Y on X

Regression Coefficient of X on Y:

- ✓ When means are given:

$$b_{xy} = \frac{\sum xy}{\sum y^2}$$

- ✓ When standard deviation are given:

$$b_{xy} = \frac{\sum xy}{n\sigma_x^2}$$

- ✓ When standard deviation and correlation of coefficient are given:

$$b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

Where $x = (X - \bar{X})$

$y = (Y - \bar{Y})$

$\sigma_x = Standard\ devitation\ of\ series\ X$

$\sigma_y = Standard\ devitation\ of\ series\ Y$

$r = Coefficent\ of\ correlation.$

Regression Coefficient of Y on X:

- ✓ When means are given:

$$b_{yx} = \frac{\sum xy}{\sum x^2}$$

- ✓ When standard deviation are given:

$$b_{yx} = \frac{\sum xy}{n\sigma_y^2}$$

- ✓ When standard deviation and correlation of coefficient are given:

$$b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

Properties of Regression Coefficients:

- ✓ The correlation coefficient is the geometric mean between two regression coefficients:
- ✓ If one of the regression coefficient is greater than one the other must be less than one so that their product can not exceed one.
- ✓ The arithmetic mean of regression coefficient is greater than r.
- ✓ The sign of correlation coefficient and the two regression coefficient will be same either all the three are positive or all the three are negative.
- ✓ Regression coefficients are independent of change of origin but they depends on change of scale. Which shows that regression coefficient are independent of change of origin but depend on change of scale.
- ✓ Both the regression lines will pass through their mean values i.e. the point $(\bar{X}, \bar{Y})$ will be on both the regression lines.

Determination of Correlation Coefficients from Regression Coefficients:

Correlation coefficient is square root of the product of two regression coefficients.

$$r = \pm\sqrt{b_{xy} \times b_{yx}}$$