MULTIPLE CORRELATION:

The problems of multiple correlation deals with situations that involves three or more variables.

Example: The association between the yield of wheat per acre and both the amount of rainfall and the average daily temperature.

We estimate the value of one of the variables based on the values of all the others.

The variables whose value we are trying to estimate is called a dependent variable and the other variables in which our estimates are based are known as independent variables.

For example, Height and age are independent variables while estimating weight of a person,

which is the dependent variable.

Coefficient of multiple correlations:

Coefficient of multiple linear correlations is represented by R and it is common to all subscripts designating the variables involved. Thus $R_{1.234}$ would represent the coefficient of multiple correlation between x1 and on one hand and X_2,$X_3$and X_4 on the other. The subscript of the dependent variable is always to the left of the point.

The coefficient of multiple correlation can be expressed in terms of r_{12}, r_{13}, r_{23} as follows.

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

$$R_{2.13}=\sqrt{\frac{r_{12}^2+r_{2.3}^2-2\,r_{12}\;r_{13}\;r_{23}}{1-r_{13}^2}}$$

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

Also $R_{2.13}$ is same $R_{2.31.}$

A coefficient of multiple correlations such as $R_{2.13}$, $R_{3.12,}$ and $R_{2.31}$ lies between 0 and 1. The value close to 1 shows the better relationship between variables. Close to ‘0’ shows worse relationship between the variable. If it is 1, the multiple correlations is perfect. If zero, no relationships exist between variables. By squaring $R_{1.23}$, we get coefficient of multiple determination.

Multiple regression equation describes the average relationship between the variables average relationship between the variables and this relationship is used to predict or control the dependent variables. An regression equation is an equation for estimating the dependent variable say X_1 from the independent variables X_2, X_3, ... and is called the regression equation of X_1 on X_2, X_3.

In case of three variables X_1, X_2, X_3. Regression equation of X_1 on X_2, X_3 has the form

$X_{1.23} = a_{1.23} + b_{12.3} X_2 + b_{13.2} X_3$

$X_{1.23}$ is the computed or estimated value of dependent variable and X_2, X_3 are the independent variables. The same equation can be represented as

$$X_1 = a_{1.23} + b_{12.3} X_2 + b_{13.2} X_3 \qquad \text{(i)}$$

If X_2 and X_3 were to be treated as dependent variables, the regression equation will be

$$X_2 = a_{2.13} + b_{21.3} X_1 + b_{23.1} X_3 \qquad \text{(ii)}$$

$$X_3 = a_{3.12} + b_{31.2} X_1 + b_{32.1} X_2 \qquad \text{(iii)}$$

The normal equations for fitting equation (i) will be

$$\Sigma X_1 = Na_{1.23} + b_{12.3} \Sigma X_2 + b_{13.2} \Sigma X_3$$
$$\Sigma X_1 . X_2 = a_{1.23} \Sigma X_2 + b_{12.3} \Sigma X_2^2 + b_{13.2} \Sigma X_2 X_3$$
$$\Sigma X_1 X_3 = a_{1.23} \Sigma X_3 + b_{12.3} \Sigma X_2 X_3 + b_{13.2} \Sigma X_3^2$$

The normal equations for fitting equation (ii) will be

$$\Sigma X_2 = Na_{2.13} + b_{21.3} \Sigma X_1 + b_{23.1} \Sigma X_3$$
$$\Sigma X_1 . X_2 = a_{2.13} \Sigma X_1 + b_{21.3} \Sigma X_1^2 + b_{23.1} \Sigma X_1 X_3$$
$$\Sigma X_2 X_3 = a_{2.13} \Sigma X_3 + b_{21.3} \Sigma X_2 X_3 + b_{23.1} \Sigma X_3^2$$

The normal equations for fitting equation (iii) will be

$$\Sigma X_3 = Na_{3.12} + b_{31.2} \Sigma X_1 + b_{32.1} \Sigma X_2$$
$$\Sigma X_1 . X_3 = a_{31.2} \Sigma X_1 + b_{31.2} \Sigma X_1^2 + b_{32.1} \Sigma X_1 X_2$$
$$\Sigma X_2 X_3 = a_{31.2} \Sigma X_3 + b_{31.2} \Sigma X_1 X_2 + b_{32.1} \Sigma X_2^2$$

1. The following zero order correlation coefficients are given: $r_{12} = 0.98$, $r_{13} = 0.44$, $r_{23} = 0.54$. Calculate multiple correlation coefficient treating first variable as dependent and the second and third variable as independent.

X_1 dependent and X_2 and X_3 are independent.

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

$$R_{1.23} = \sqrt{\frac{(0.98)^2 + (0.44)^2 - 2 \times 0.98 \times 0.44 \times 0.54}{1 - 0.54^2}}$$

$$= \sqrt{\frac{0.9604 + 0.1936 - 0.4657}{1 - 0.7084}} = 0.986$$

2. Find the multiple linear regression equation of X_1 on X_2 and X_3 from the following to three variables given below:

X_1	4	6	7	9	13	15
X_2	15	12	8	6	4	3
X_3	30	24	20	14	10	4

The regression equation of X_1 on X_2 and X_3 is

$X_1 = a_{1.23} + b_{12.3} X_2 + b_{13.2} X_3$

The value of the constants $a_{1.23}$, $b_{12.3}$, $b_{13.2}$ is obtained by solving the following set of equations (Normal Equations)

$$\sum X_1 = N a_{1.23} + b_{12.3} \sum X_2 + b_{13.2} \sum X_3$$
$$\sum X_1 X_2 = a_{1.23} \sum X_2 + b_{12.3} \sum X_2^2 + b_{13.2} \sum X_2 X_3$$
$$\sum X_1 X_3 = a_{1.23} \sum X_3 + b_{12.3} \sum X_2 X_3 + b_{13.2} \sum X_3^2$$

Then calculating the required values:

X_1	X_2	X_3	X_1X_3	$X_1 X_2$	$X_2 X_3$	X_1^2	X_2^2	X_3^2
4	15	30	120	60	450	16	225	900
6	12	24	144	72	288	36	144	576
7	8	20	140	56	160	49	64	400
9	6	14	126	54	84	81	36	196
13	4	10	130	52	40	169	16	100
15	3	4	60	60	12	225	9	16
$\Sigma X_1 = 54$	$\Sigma X_2 = 48$	$\Sigma X_3 = 102$	$\Sigma X_1 X_3 = 720$	$\Sigma X_1 X_2 = 339$	$\Sigma X_2 X_3 = 1034$	$\Sigma X_1^2 = 576$	$\Sigma X_2^2 = 494$	$\Sigma X_3^2 = 2188$

Substituting the values in the normal equations:

$$6a_{1.23} + 48b_{12.3} + 102b_{13.2} = 54 \qquad (i)$$

$$48a_{1.23} + 49b_{12.3} + 1034b_{13.2} = 339 \qquad (ii)$$

$$102a_{1.23} + 1034b_{12.3} + 2188b_{13.2} = 720 \qquad (iii)$$

Solving (i) , (ii) and (iii), we get

$a_{1.23} = 16.479$, $b_{12.3} = 0.389$ and $b_{13.2} = - 0.623$

$X_1 = 16.479 + 0.389\ X_2 - 0.623\ X_3$