Definition: Multiple regression analysis is a statistical method used to predict the value a dependent variable based on the values of two or more independent variables.

What Does Multiple Regression Analysis Mean?
What is the definition of multiple regression analysis? The value being predicted is termed dependent variable because its outcome or value depends on the behavior of other variables. The independent variables' value is usually ascertained from the population or sample.
In business, sales managers use multiple regression analysis to analyze the impact of some promotional activities on sales. Multiple regression analysis can be used to also unearth the impact of salary increment and increments in other employee benefits on employee output. The analysis is useful when you want to predict the impact of individual independent variables on the desired outcome.

This analysis makes some assumptions on the margin of error for the analysis, which needs to be checked when using the model. The most common is that, the errors are independent and normally distributed. It also assumes the errors have constant variance and the mean of the errors is zero.

Let's look at an example.

Example
Multiple regression analysis can be performed using Microsoft Excel and IBM's SPSS. Other statistical tools can equally be used to easily predict the outcome of a dependent variable from the behavior of two or more independent variables.

This analysis can be used to predict how well a new process in a company is responding to some tweaks made to that process. It can also be used at home to ascertain changes in the cost of energy consumed based on some energy conservation methods and equipment employed in the house. In schools, this analysis is used to determine the performance of students using class hours, library hours, and leisure hours as the independent variables.Summary Definition

Define Multiple Regression Analysis: MRA means a method of predicting outcomes based on manipulating one variable at a time.
We have

$$Y = a + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k + e$$

$$Y' = a + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k$$

Basically the reason for using these formula for the calculation regression coefficients b1 and b2 in our case is just the formula in b1 and b2 is basically the varaince to x1 n x2 so we cannot use them directly as given in the formula for example $\sum x1^2$ is basicallly used in lower case n this specifies like this is variance to x1^2 so thats y for the calculation of $\sum x1^2$, *we use* $\sum x1^2 = \sum X1X2 - \frac{\sum X1 \cdot \sum X2}{N}$

...... SO ON

Variable selection: we have to always focuss on the p-value generated by the software and we must notice if the values >= 0.15 then don't consider the predictor or explanatory variable as part of ur model

$$\sum x_1^2 = \sum X_1X_1 - \frac{(\sum X_1)(\sum X_1)}{N}$$

$$\sum x_2^2 = \sum X_2X_2 - \frac{(\sum X_2)(\sum X_2)}{N}$$

$$\sum x_1 y = \sum X_1 Y - \frac{(\sum X_1)(\sum Y)}{N}$$

$$\sum x_2 y = \sum X_2 Y - \frac{(\sum X_2)(\sum Y)}{N}$$

$$\sum x_1 x_2 = \sum X_1 X_2 - \frac{(\sum X_1)(\sum X_2)}{N}$$

the two variables that is it shows the average probable change in one variable given a certain change in another variable.

Regression line X on Y

$$X = c + dy$$

where

$$b = d = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum y^2 - (\sum y)^2}$$

$$a = c = \bar{x} - d\bar{y}$$

X	1	5	3	2	1	1	7	3	2.875	$\bar{x}$
Y	6	1	0	0	1	2	1	5	2	$\bar{y}$

a) Fit the Regression line of Y on X and hence predict Y if X = 10.

b) Fit the Regression line of X on Y and hence predict X; if Y = 2.5

a) $\boxed{2.8745 - 0.3042x}$

b) $\boxed{3.4306 - 0.2778Y}$

Hint: Also focus largest standard error of estimate $\sigma_Y = \sqrt{\frac{\sum(Y-\hat{Y})^2}{N}}$ will show the poorer value of correlation.

Correlation: specifies degree and strengthening of relationship where as regression always show the nature of relationship b/w the we have

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

so $R_{1y} = r_{y1} = \dfrac{5(95.8) - (20)(19.5)}{\sqrt{[5(90) - (20)^2][5(196.93) - (19.5)^2]}}$

$= \boxed{0.511967}$

o) $R_{2y} = r_{y_2} = \dfrac{5(45.6) - (24)(19.5)}{\sqrt{[5(148) - (24)^2][5(196.93) - (19.5)^2]}}$

$= -0.7623$

$R_{12} = r_{12} = \dfrac{5(99) - (20)(24)}{\sqrt{[5(90) - (20)^2][5(148) - (24)^2]}}$

$= 0.1656$

R

R^2 = adjusted R^2

Prediction Analysis (model)

Multiple Regression

Y	X_1	X_2	X_1^2	X_2^2	$X_1 \times X_2$	X_1Y	X_2Y	Y^2
−3.7	3	8	9	64	24	−11.1	−30	13.69
3.5	4	5	16	25	20	14	17.5	12.25
2.5	5	7	25	49	35	12.5	17.5	6.25
11.5	6	3	36	9	18	69	34.5	132.25
5.7	2	1	4	1	2	11.4	5.7	32.49
Σ19.5	20	24	90	148	99	95.8	15.6	196.93
3.9	4	4.8						

Substituting the values in we get

$$Y = b_0 + b_1X_1 + b_2X_2$$

$$Y = 2.7996 + 2.2816X_1 - 1.6721X_2$$

for correlation in multiple regression, we need to first find and observe

$$R = r^2_{Y_1} + r^2_{Y_2} \text{ (only if } r_{12} = 0)$$

Otherwise

$$R = \frac{r^2_{Y_1} + r^2_{Y_2} - 2r_{Y_1}r_{Y_2}r_{12}}{1 - r^2_{12}}$$

$R = ?$ $\boxed{0.9998}$

We have

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

So

$$R_{1Y} = r_{Y_1} = \frac{5(95.8) - (20)(19.5)}{\sqrt{[5(90) - (20)^2][5(196.93) - (19.5)^2]}}$$

$$= \boxed{0.511967}$$

$$R_{2Y} = r_{Y_2} = \frac{5(45.6) - (24)(19.5)}{\sqrt{[5(148) - (24)^2][5(196.93) - (19.5)^2]}}$$

$$= -0.7623$$

$$R_{12} = r_{12} = \frac{5(99) - (20)(24)}{\sqrt{[5(90) - (20)^2][5(148) - (24)^2]}}$$

$$= 0.1656$$

	y	x1	x2	x1^2	x2^2	x1*x2	x1y	x2y	y^2
	-3.7	3	8	9	64	24	-11.1	-30	13.69
	3.5	4	5	16	25	20	14	17.5	12.25
	2.5	5	7	25	49	35	12.5	17.5	6.25
	11.5	6	3	36	9	18	69	34.5	132.25
	5.7	2	1	4	1	2	11.4	5.7	32.49
sum	19.5	20	24	90	148	99	95.8	45.6	$\boxed{196.93}$
MEAN	3.9	4	4.8						

We have

a =			a=	2.7996	b1	2.2816	b2	-1.6721

$$\overline{a} = \overline{Y} - b_1\overline{X1} - b_2\overline{X2}$$

$$b_1 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$b_2 = \frac{(\sum x_1^2)(\sum x_2 y) - (\sum x_1 x_2)(\sum x_1 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$R^2 = r_{y1}^2 + r_{y2}^2 \quad (only\ if\ r_{12} = 0)$$

$$R^2 = \frac{r_{y1}^2 + r_{y2}^2 - 2r_{y1}r_{y2}r_{12}}{1 - r_{12}^2}$$

square values g Ry

R1Y	0.511968	0.262111
R2Y	-0.7623	0.581104
R12	0.165647	0.027439

R12 is a correlation between x1 n x2

sx1^2	10
sx2^2	32.8
sx1y	17.8
sx2y	-48
sx1x2	3

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

$$R_{1y} = r_{1y} = \frac{5(95.8) - (20)(19.5)}{\sqrt{[5(90) - (20)^2][5(196.93) - (19.5)^2]}} = \underline{0.511967}$$

$$R_{2y} = r_{2y} = \frac{5(45.6) - (24)(19.5)}{\sqrt{[5(148) - (24)^2][5(196.93) - (19.5)^2]}} = -0.7623$$

$$R_{12} = r_{12} = \frac{5(99) - (20)(24)}{\sqrt{[5(90) - (20)^2][5(148) - (24)^2]}} = 0.1656$$

$$R = \frac{r_{y_1}^2 + r_{y_2}^2 - 2r_{y_1} \cdot r_{y_2} \cdot r_{12}}{1 - r_{12}^2}$$

$$R = \frac{0.2621 + 0.5811 - 2(0.5119)(-0.7623)(0.1656)}{1 - (0.1656)^2}$$

$$\boxed{R = 0.9998}$$ (coefficient of determination).

Adjusted R^2

$$R^2_{adj} = 1 - \frac{(1-R^2)(n-1)}{n-k-1}$$

where k = no. of independent variables.

$$R^2_{adj} = 1 - \frac{(1-0.998^2)(5-1)}{5-2-1} = \boxed{0.9920}$$