Business Statistics

Dr.M.K.Singh **BBA SEM II**

Regression Analysis

Regression Analysis is Concerned with finding cause and effect relationship between two or more than two variables. If there exist a relation ship between two variables Then the change in one variable may be due to the change in other variable that is , one may be the cause of change and the other may be the effect. For example ,if there is a change in income then there will be a change in expenditure or saving. In this case change in income will be the cause of change in expenditure and change in expenditure will be the effect.

In business, several times it becomes necessary to have some forecast so that the management can take a decision regarding a product or a particular course of action. In order to make a forecast, one has to ascertain some relationship between two or more variables relevant to a particular situation. For example, a company is interested to know how far the demand for television sets will increase in the next five years, keeping in mind the growth of population in a certain town. Here, it clearly assumes that the increase in population will lead to an increased demand for television sets. Thus, to determine the nature and extent of relationship between these two variables becomes important for the company.

Simple regression involves only two variables; one variable is predicted by another variable. The variable to be predicted is called the **dependent variable**. The predictor is called the **independent variable,** or explanatory variable. For example, when we are trying to predict the demand for television sets on the basis of population growth, we are using the demand for television sets as the dependent variable and the population growth as the independent or predictor variable.

The cause and effect relationship between two variables are expressed by two regression equations :

X= a+bY ---(1) Regression Equation for X on Y

Y =a+bX ---(2) Regression Equation for Y on X

Where, a= constants
and b= regression coefficients

The problem is to find the values of a and b.The method of finding the values of a and b is known as The Least Square Method.In the method of least squares the values of *a* and *b* are obtained by solving simultaneously the following pair of normal equations.After obtaining the values of a and b the regression equation for Y on X can be obtained as Y= a+bX .

$\sum Y = aN + b\sum X$ ....(1)

$\sum XY = a\sum X + b\sum X^2$ ....(2)

X	Y	X^2	Y^2	XY
14	10	196	100	140
17	12	289	144	204
23	15	529	225	345
21	20	441	400	420
25	23	625	529	575
100	80	2080	1398	1684

Substituting the values in equation (1) and equation (2) we get,

80 =5a +100b ...(1)
1648=100a +2080b ...(2)

Solving these simultaneous equations for a and b we get,

b=1.05 and a=-5 ,so the equation for Y on X will be,

Y=-5+1.05X

Regression Equation for X on Y

Regression equation for X on Y can be obtained from solving the following pairs of normal equations :

$\sum X = aN + b\sum Y$ ....(1)

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

Substituting the Values ,we get,

100 =5a+80b .....(1)
1684=80a+1398b .....(2)

Solving these simultaneous equations we get,

a=8.608 and b=0.712

Thus the regression equation for X on Y will be,

X = 8.608 +0.712 Y

Regression Coefficients

Let us consider the line of regression of Y on X,

Y =a+bX

where b which is the slope of the line of regression of Y on X is called the coefficient of regression of y on x. It represents the increment in the value of the dependent variable Y for a unit change in the value of independent variable x.For notational convenience the coefficient of regression of y on x is written as b_{yx} and the coefficient of regression of x on y is written as b_{xy},that is ,

b_{yx}=regression coefficient of y on x

b_{xy}=regression coefficient of x on y

The coefficient of regression of y on x is given by

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

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

where, r = coefficient of correlation

b_{yx} = regression coefficient of y on x,and
b_{xy}= regression coefficient of x on y .

The value of constant (parameter) a for regression equation y on x can be obtained as

$a(yx)=\bar{Y} - b_{yx}*\bar{X}$ and for equation x on y can be obtained as,

$a(xy)= \bar{X} - b_{xy}*\bar{Y}$

Ex.
Given,
$\bar{X}=40$, $\bar{Y}=6$, $\sigma_x=10$, $\sigma= 1.5$ and $r=+0.9$
Find regression equations for y on x and x on y .

Solution:

Regression Equation for y on x

$Y=a+bx \quad =0.6+0.135x$

Regression equation for X on Y

$X= a+bY \quad = 4+6Y$

$r= \sqrt{bxy*byx} =\sqrt{0.135*6} =\sqrt{0.81} =0.9$

Exercise 1

You are given the following data:

	X	Y
Arithmetic mean	36	85
Standard deviation	11	8
Correlation coefficient	0.66	

(i) Find the two regression equations
(ii) Estimate value of x when y=32
Ans.(i)y=0.48x+67.72 , x=0.9075y-41.1375 ; (ii) 26.925

Exercise 2.
You are given the following information about advertisement expenditure and sales :

	Advertisement expenditure (x)	Sales(Y)
Mean	20	120
S. D	5	25
Correlation coefficient	0.8	

(i) Calculate the two regression equations
(ii) Find the likely sales when advertisement expenditure is Rs.25 crore.
(iii) What should be the advertisement budget if the company wants to attain the sales target of Rs 159 crore?
Ans. (i) X=0.16Y+0.8 ,Y=4X +4
(iii) Rs.140 crore ,(iii) Rs.24.8 crore

Exercise 3
From the following data find the probable yield when the rainfall is 29”

	Rainfall	Yield
Mean	25”	40 units/hectare
Standard deviation	3”	6 unit/hectare

Correlation coefficient = 0.8
Ans. 33.6 unit/hectare