Regression Analysis

Dr. Sabiha Khan

Meaning of Regression

The dictionary meaning of the word Regression is ***'Stepping back' or 'Going back'***. Regression is the measures of the average relationship between two or more variables in terms of the original units of the data. And it is also attempts to establish the nature of the relationship between variables that is to study the functional relationship between the variables and thereby provide a mechanism for prediction, or forecasting.

Regression Analysis

- The statistical technique of estimating the unknown value of one variable (i.e., ***dependent variable***) from the known value of other variable (i.e., ***independent variable***) is called regression analysis.
- How the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed.

Examples:

- The effect of a price increase upon demand, for example, or the effect of changes in the money supply upon the inflation rate.
- Factors that are associated with variations in earnings across individuals—occupation, age, experience, educational attainment, motivation, and ability. For the time being, let us restrict attention to a single factor—call it education. Regression analysis with a single explanatory variable is termed "simple regression."

Importance of Regression Analysis

Regression analysis helps in three important ways :-

1. It provides estimate of values of dependent variables from values of independent variables.

2. It can be extended to 2or more variables, which is known as multiple regression.

3. It shows the nature of relationship between two or more variable.

Simple Linear Regression Model

- The fundamental aim or regression analysis is to determine a regression equation (line) that makes sense and fits the representative data such that the error of variance is as small as possible.

A regression line is a line that best describes the linear relationship between the two variables. It is expressed by means of an equation of the form:

The Regression equation of ***X on Y*** is:

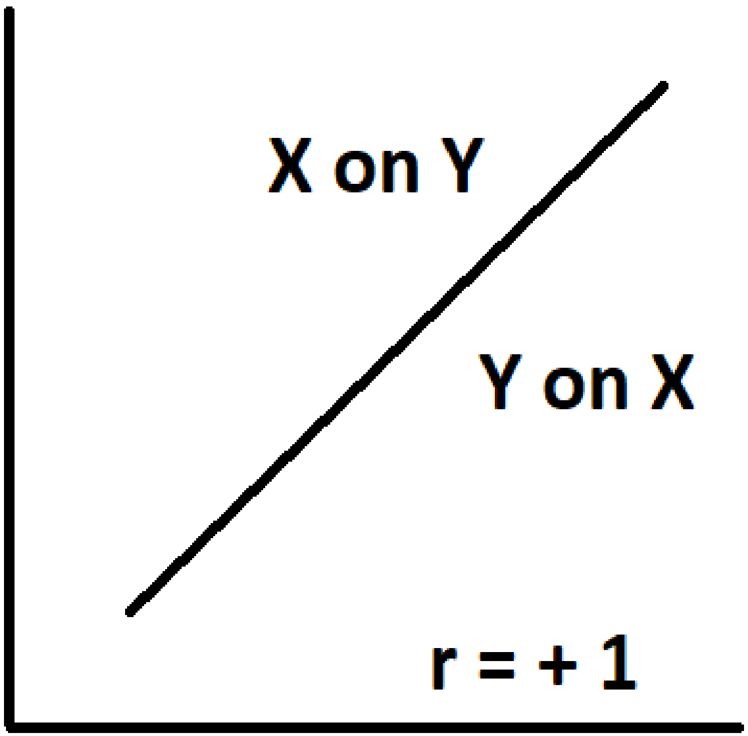
$$X = a + bY$$

The Regression equation of ***Y on X*** is:

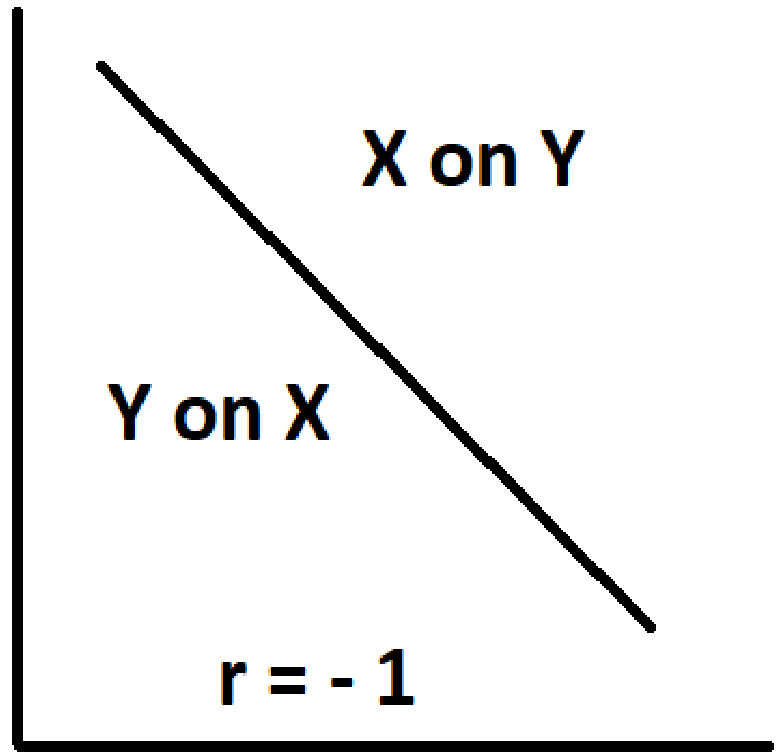
$$Y = a + bX$$

Regression Lines and Coefficient of Correlation

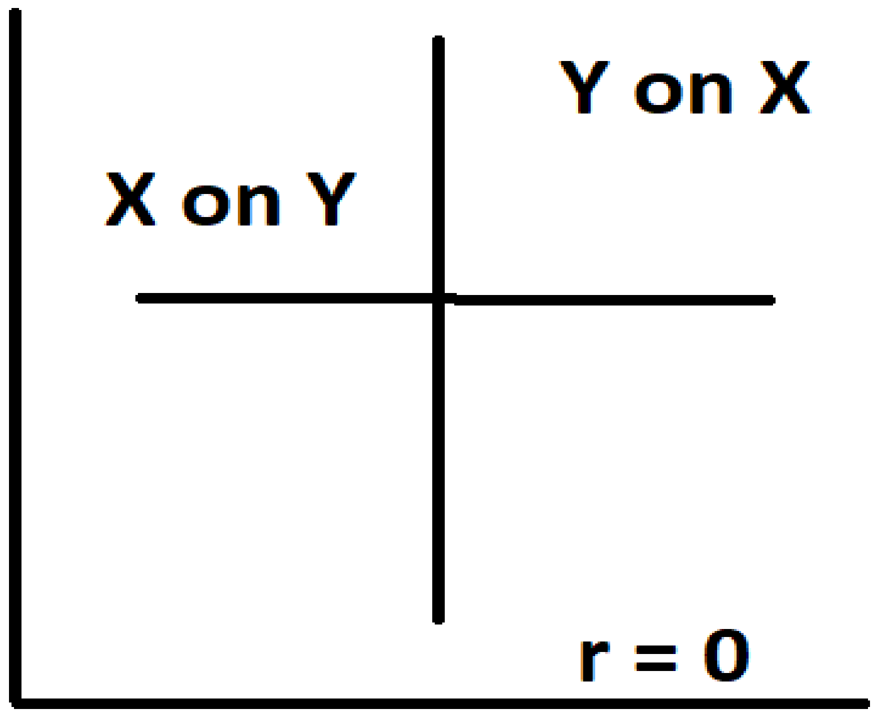
Perfect Positive Correlation



Perfect Negative Correlation



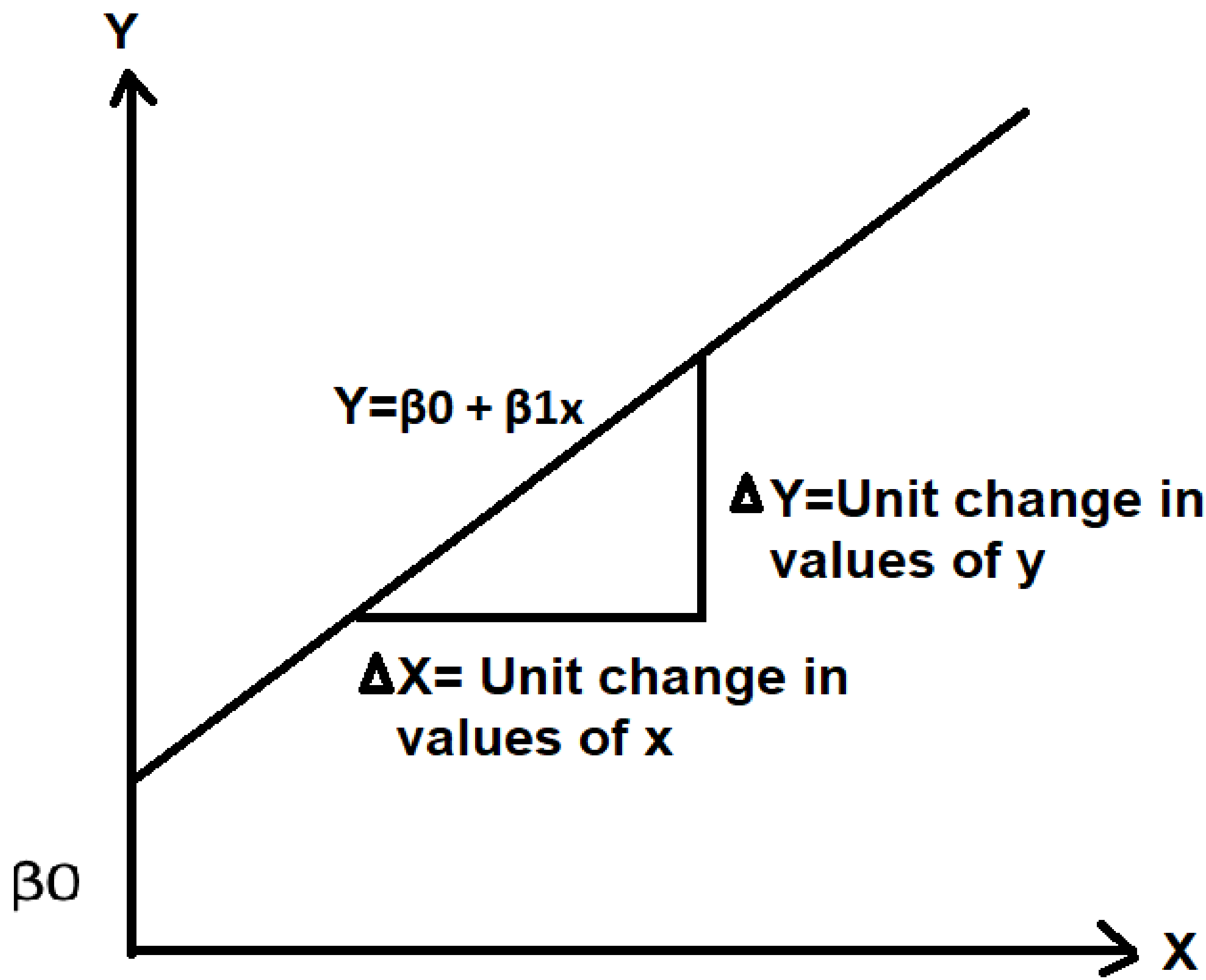
No Correlation



Regression Coefficients

- It is difficult to determine the values of population parameter β0 and β1 in equation y = β0 + β1x, because we cannot observe the entire population.
  - Here,
  - β0 = Average value of dependent variable y when x = 0
  - B1 = Expected change in the value of y for a unit change in the value of x.
- This equation requires the determination of two population coefficients β0 and β1 in order to predict the average value of y.

Straight Line Relationship



- The sample equation representing the straigth line regression model can be written as:
- $y = a + bx$
  - Where,
  - y = estimated average (mean) value of dependent variable y for a given value of independent variable x.
  - a = y-intercept that represents average value of y.
  - b = slope of regression line that represents the expected change in the value of y for unit change in the value of x.
- To determine the value of y for a given value of x, this equation requires the determination of two unknown constants a) and b(regression coefficient) .
  - a = (intercept or represents the level of fitted line or th distance of the line above or below the origin, when x=0.
  - b = regression coefficient or represents the slope of the line or a measure of the change in the estimated value of y for a one –unit change in x.

Methods to determine the parameters of a regression equation:
1. Method of Normal Equation.

The regression equation of X on Y is :

$$X = a + bY$$

Where,

X=Dependent variable

Y=Independent variable

The regression equation of Y on X is:

$$Y = a + bX$$

Where,

Y=Dependent variable

X=Independent variable

And the values of a and b in the above equations are found by the method of least of Squares-reference . The values of a and b are found with the help of normal equations given below:

(I)

$$\sum X = na + b\sum Y$$
$$\sum XY = a\sum Y + b\sum Y^2$$

(II)

$$\sum Y = na + b\sum X$$
$$\sum XY = a\sum X + b\sum X^2$$

Example1-: From the following data obtain the two regression equations using the method of Least Squares.

X	3	2	7	4	8
Y	6	1	8	5	9

Solution-:

X	Y	XY	X^2	Y^2
3	6	18	9	36
2	1	2	4	1
7	8	56	49	64
4	5	20	16	25
8	9	72	64	81
$\sum X = 24$	$\sum Y = 29$	$\sum XY = 168$	$\sum X^2 = 142$	$\sum Y^2 = 207$

$$\sum Y = na + b\sum X \qquad \sum XY = a\sum X + b\sum X^2$$

Substitution the values from the table we get

29=5a+24b…………………(i)
168=24a+142b
84=12a+71b………………...(ii)

Multiplying equation (i) by 12 and (ii) by 5

348=60a+288b………………(iii)
420=60a+355b………………(iv)

By solving equation(iii)and (iv) we get

a=0.66 and b=1.07

By putting the value of a and b in the Regression equation Y on X we get

Y=0.66+1.07X

Normal equation Method

$\Sigma X = na + b\Sigma Y$ and $\Sigma XY = a\Sigma Y + b\Sigma Y^2$

$\Sigma Y = na + b\Sigma X$ and $\Sigma XY = a\Sigma X + b\Sigma X^2$

$\Sigma Y = na + b\Sigma X$ — (1)

समीकरण (1) में मूल्य रखने पर

$29 = 5a + 24b$

$\Sigma XY = a\Sigma X + b\Sigma X^2$ — (2)

समीकरण (2) में मूल्य रखने पर

$168 = 24a + 142b$

समीकरण को सरलीकृत करने पर

$168 = 2(12a + 71b)$ or $84 = 12a + 71b$

$\frac{168}{2} = 12a + 71b$

समीकरण (1) को 12 से तथा समीकरण (2) को 5 से गुणा करने पर

$(29 = 5a + 24b) \times 12$ — (3)

$(84 = 12a + 71b) \times 5$ — (4)

हल करने

$348 = 60a + 288b$ — (3)

$\underline{420 = 60a + 355b}$ — (4)

$-72 = -67b$

$b = \frac{+72}{+67} = 1.07$

b का मान समीकरण (1) में रखने पर

$29 = 5a + 24b$

$29 = 5a + 24 \times 1.07$

$29 = 5a + 25.68$

$29 = 5a + 25.68$

$5a = 29 - 25.68$

$a = \frac{3.32}{5} = 0.66$

Regression equation Y on x = $\boxed{Y = a + bX}$

इस समीकरण मे a तथा b का मान रखने पर

$Y = 0.66 + 1.07x$

इसी प्रकार Regression equation x on y = $\boxed{x = a + by}$

$\Sigma x = na + b\Sigma y$ — (1)

$\Sigma xy = a\Sigma y = b\Sigma y^2$ — (2)

समीकरण (1) में मान रखने पर

$24 = 5a + 29b$ — (1)

समीकरण (2) में मान रखने पर

$168 = 29a + 207b$ — (2)

समी॰ (1) को 29 से तथा समी॰ (2) को 5 से गुणा करने पर

$(24 = 5a + 29b) \times 29$ — (3)

$(168 = 29a + 207b) \times 5$ — (4)

समीकरण को हल करने पर -

$696 = 145a + 841b$ — (3)

$\underline{840 = 145a + 1035b}$ — (4)

$-144 = -194b$

$b = \frac{144}{194} = 0.74$

b का मान समीकरण (1) में रखने पर

$24 = 5a + 29b$

$24 = 5a + 29 \times 0.74$

$24 = 5a + 21.46$

$\Rightarrow 5a = 24 - 21.46$

$a = \frac{2.54}{5} = 0.50$

a और b का समीकरण में रखने पर

$x = a + bY$

$x = 0.50 + 0.74Y$

Now to find the regression equation of X on Y ,
The two normal equation are

$$\sum X = na + b\sum Y$$
$$\sum XY = a\sum Y + b\sum Y^2$$

Substituting the values in the equations we get

24=5a+29b……………………….(i)
168=29a+207b…………………..(ii)

Multiplying equation (i)by 29 and in (ii) by 5 we get

a=0.49 and b=0.74

Substituting the values of a and b in the Regression equation X and Y

X=0.49+0.74Y

2. Deviation Method (By Actual Mean)

Deviations from the Arithmetic mean method:

The calculation by the least squares method are quit difficult when the values of X and Y are large. So the work can be simplified by using this method.
The formula for the calculation of Regression Equations by this method:

Regression Equation of X on Y-

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

Regression Equation of Y on X-

$$(Y-\bar{Y})=b_{yx}(X-\bar{X})$$

Where, b_{xy} and b_{yx} = Regression Coefficient

$$b_{xy}=\frac{\sum xy}{\sum y^2} \quad \text{and} \quad b_{yx}=\frac{\sum xy}{\sum x^2}$$

Example2-: From the previous data obtain the regression equations by Taking deviations from the actual means of X and Y series.

X	3	2	7	4	8
Y	6	1	8	5	9

Solution-:

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	y^2	xy
3	6	-1.8	0.2	3.24	0.04	-0.36
2	1	-2.8	-4.8	7.84	23.04	13.44
7	8	2.2	2.2	4.84	4.84	4.84
4	5	-0.8	-0.8	0.64	0.64	0.64
8	9	3.2	3.2	10.24	10.24	10.24
$\sum X = 24$	$\sum Y = 29$	$\sum x = 0$	$\sum y = 0$	$\sum x^2 = 26.8$	$\sum y^2 = 38.8$	$\sum xy = 28.8$

Regression Equation of X on Y is

$$(X-\overline{X})=b_{xy}(Y-\overline{Y})$$

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

$$X-4.8=\frac{28.8}{38.8}(Y-5.8)$$

$$X-4.8=0.74(Y-5.8)$$

$$X=0.49+0.74Y \quad \text{............(I)}$$

Regression Equation of Y on X is

$$(Y-\overline{Y})=b_{yx}(X-\overline{X})$$

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

$$Y-5.8=\frac{28.8}{26.8}(X-4.8)$$

$$Y-5.8=1.07(X-4.8)$$

$$Y=0.66+1.07X \quad \text{............(II)}$$

It would be observed that these regression equations are same as those obtained by the direct method .

Deviation from Assumed mean method-:

When actual mean of X and Y variables are in fractions ,the calculations can be simplified by taking the deviations from the assumed mean.

The Regression Equation of X on Y-:

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

The Regression Equation of Y on X-:

$$(Y-\bar{Y})=b_{yx}(X-\bar{X})$$

But , here the values of b_{xy} and b_{yx} will be calculated by following formula:

$$b_{xy}=\frac{N\sum d_x d_y-\sum d_x \sum d_y}{N\sum d_y^2-\left(\sum d_y\right)^2} \qquad b_{yx}=\frac{N\sum d_x d_y-\sum d_x \sum d_y}{N\sum d_x^2-\left(\sum d_x\right)^2}$$

Example-3: From the data given in previous example calculate regression equations by assuming 7 as the mean of X series and 6 as the mean of Y series.

Solution-:

X	Y	Dev. From assu. Mean 7 $(d_x)=X-7$	d_x^2	Dev. From assu. Mean 6 $(d_y)=Y-6$	d_y^2	d_xd_y
3	6	-4	16	0	0	0
2	1	-5	25	-5	25	+25
7	8	0	0	2	4	0
4	5	-3	9	-1	1	+3
8	9	1	1	3	9	+3
$\sum X=24$	$\sum Y=29$	$\sum d_x=-11$	$\sum d_x^2=51$	$\sum d_y=-1$	$\sum d_y^2=39$	$\sum d_xd_y=31$

$$\bar{X} = \frac{\sum X}{N} \Rightarrow \bar{X} = \frac{24}{5} = 4.8 \qquad \bar{Y} = \frac{\sum Y}{N} \Rightarrow \bar{Y} = \frac{29}{5} = 5.8$$

The Regression Coefficient of X on Y-:

$$b_{xy} = \frac{N\sum d_x d_y - \sum d_x \sum d_y}{N\sum d_y^2 - \left(\sum d_y\right)^2}$$

$$b_{xy} = \frac{5(31) - (-11)(-1)}{5(39) - (-1)^2}$$

$$b_{xy} = \frac{155 - 11}{195 - 1}$$

$$b_{xy} = \frac{144}{194}$$

$$b_{xy} = 0.74$$

The Regression equation of X on Y-:

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

$$(X - 4.8) = 0.74(Y - 5.8)$$

$$X = 0.74Y + 0.49$$

The Regression coefficient of Y on X-:

$$b_{yx} = \frac{N\sum d_x d_y - \sum d_x \sum d_y}{N\sum d_x^2 - \left(\sum d_x\right)^2}$$

$$b_{yx} = \frac{5(31) - (-11)(-1)}{5(51) - (-11)^2}$$

$$b_{yx} = \frac{155 - 11}{255 - 121}$$

$$b_{yx} = \frac{144}{134}$$

$$b_{yx} = 1.07$$

The Regression Equation of Y on X-:

$$(Y - \bar{Y}) = b_{yx}(X - \bar{X})$$

$$(Y - 5.8) = 1.07(X - 4.8)$$

$$Y = 1.07X + 0.66$$

It would be observed the these regression equations are same as those obtained by the least squares method and deviation from arithmetic mean .

Question.1: Use least squares regression lines to estimate the increase in sales revenue expected from an increase of 7.5 per cent in advertising expenditure.

FIRM	Annual percentage increase in advertising expenditure	Annual percentage increase in sales revenue
A	1	1
B	3	2
C	4	2
D	6	4
E	8	6
F	9	8
G	11	8
H	14	9

Question: 2 : The following data gives the age and blood pressure of 10 women.

Age	**56**	**42**	**36**	**47**	**49**	**42**	**60**	**72**	**63**	**55**
Blood Pressure	147	125	118	128	145	140	155	160	149	150

- a) Find the correlation coefficient between age and blood pressure.
- b) Determine the least square regression equation of blood pressure on age.
- c) Estimate the blood pressure of a woman whose age is 45 years.

Example 3: The General Sales Manager of Kiran Enterprises-an enterprise dealing in the sale of readymade men's wear-is toying with the idea of increasing his sales to Rs 80,000. On checking the records of sales during the last 10 years, it was found that the annual sale proceeds and advertisement expenditures were highly correlated to the extent of 0.8 . It was further noted that the annual average sale has been Rs 45,000 and annual average sale has been Rs 45,000 and annual average advertisement expenditure Rs 30,000 , with a variance of Rs 1600 and Rs 625 in advertisement expenditure respectively.

In view of the above, how much expenditure on advertisement would you suggest the General Sales Manager of the enterprise to incur to meet his target of sales?

THANK YOU