Regression Analysis

Regression Analysis is measure of the average relationship between two or more variables in terms of the original units of the data. One of these variables is called the dependent variable and another variable is known as independent variable.

Thus, the determination of an appropriate functional relationship between the variables is termed as the 'Regression analysis'.

The statistical technique of estimating or predicting the unknown value of a dependent variable from known value of an independent variable is called Regression Analysis.

Objectives of Regression Analysis:

1 Forecasting or Prediction
2. Helpful in planning & decision making
3. Testing of Economic Theory
4. Useful in Economics & Commerce or social

ASSUMPTIONS OF REGRESSION ANALYSIS

Regression analysis is based on the following assumptions :

1. Regression analysis assumes linear relationship between related variables. It means the relationship can be expressed by a straight line $Y = a + bX$ or $X = a' + b'Y$.
2. There is no error of measurement and aggregation error in the variable.
3. The relationship under study is exactly identified.
4. The residual errors are normally distributed. They have zero mean and constant variance.

COMPARISON BETWEEN CORRELATION AND REGRESSION

Although correlation and Regression are both concerned with the relationship of two or more variables but we have some basic differences in the concept of these two which are as under ;

Correlation	Regression
(*i*) It studies the co-variability between two variables. It tells whether the two variables move in the same or in the opposite direction.	(*i*) It expresses the average relationship between two or more variables.
(*ii*) Correlation does not establish cause and effect relationship between the variables.	(*ii*) Regression analysis is based on cause and effect relationship between the variables. The variables expressing cause are taken as independent variables and that expressing effect is taken as dependent variable.
(*iii*) If there is a mathematical or statistical relationship between the variables, it is numerically determinant and expressed by *r*. But in case of illogical relationship the correlation is known as non-sense or spurious correlation.	(*iii*) Regression analysis regarding the relationship between X and Y can never be non-sense. For every X we get some value of Y and for every Y we get some value of X.
(*iv*) Correlation coefficient between the variables X and Y or Y and X is always symmetrical. It means that , $r_{xy} = r_{yx}$	(*iv*) Regression coefficients may or may not have the property of symmetry. It means that, b_{xy} and b_{yx} may or may not be equal.
(*v*) Coefficient of correlation always lies between – 1 and + 1.	(*v*) Regression coefficients can have any value but the product of two regression coefficients is always positive and never exceed one. It means, $0 \leq (bxy) \times (byx) \leq 1$
(*vi*) Coefficient of correlation is independent of change of origin and scale.	(*vi*) Regression coefficients are independent of change of origin but not of scale.

Regressive Analysis

The statistical technique of estimating or predicting the unknown value of a dependent variable from the known value of an Independent variable is called Regression Analysis

Calculating Regression equation.
or Derivation of Regression lines

i) Regression Equation through Normal Equation.

ii) Regression Equation through Regression Co-efficient

1) Regression Equation through Normal Equation

Regression line X on Y

$X_c = a + bY$

$\Sigma X = Na + b\Sigma Y$ — ①

$\Sigma XY = a\Sigma Y + b\Sigma Y^2$ — ②

Regression line Y on X

$Y_c = a + bX$

$\Sigma Y = Na + b\Sigma X$ — ①

$\Sigma XY = a\Sigma X + b\Sigma X^2$ ②

c is constant

Q) Given the Bivariate data

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

i) Fit the regression line X on Y
predict X if Y = 5

classmate

Date ______

Page ______

Fit the Regression line Y on X if X = 20

Sol

X	Y	X^2	Y^2	XY
2	7	4	49	14
6	2	36	4	12
4	1	16	1	4
3	1	9	1	3
2	2	4	4	4
2	3	4	9	6
8	2	64	4	16
4	6	16	36	24
$\Sigma x = 31$	$\Sigma Y = 24$	153	108	$\Sigma xy = 83$

The Regression equation of X on Y is

$$X = a + bY$$

The two normal equation are

$$\Sigma X = Na + b\Sigma Y$$

$$\Sigma XY = a\Sigma Y + b\Sigma Y^2$$

Substitute the value, we get

$$31 = 8a + 24b \quad \text{--- (1)}$$

$$83 = 24a + 108b \quad \text{--- (11)}$$

Multiply eq (1) by 3 & eq 2 by 1, we get

$$93 = 24a + 72b$$

$$83 = 24a + 108b$$

$$10 = -36b$$

or $b = -\frac{10}{36} = -0.28$

put the value of b eq (1)

$$31 = 8a + 24(-0.28)$$

$$31 = 8a - 37.72$$

$$8a = 37.72$$

$$a = \frac{37.72}{8} = 4.715$$

Scanned with CamScanner

classmate
Date
Page

Regression equation of X & Y is
$X = 4.715 - 0.28y$
Value of X when Y = 5
or $X = 4.715 - \cancel{0.28}\ 0.28(5)$
or $X = 4.715 - 1.40$
or $X = 3.315$

(i) Regression line Y on X
$Y = a + bx$
The two normal equations are
$\Sigma y = Na + b\Sigma x$
$\Sigma xy = a\Sigma x + b\Sigma x^2$
Substitute the values, we get
$24 = 8a + b(31)$ — (1)
$83 = a(31) + b(153)$ — (2)
Multiplying equations (1) by 31 & eq. (2) by 8

$744 = 248a + 961b$
$\underline{- 644 = 248a + 1224b}$
$80 = 263b$
$b = -80/263 = -0.304$
put the value of b in eq. (1), we get
$24 = 8a + 31(-0.304)$
$24 = 8a - 9.424$
$24 + 9.424 = 8a$
$33.3 = 8a$
$a = \frac{33.3}{8} = 4.162$

$Y = a + bX$
$y = 4.162 +$

Regression Equations through Regression Coefficients
taking Deviations from Actual mean

x on y

$X - \bar{X} = bxy\,(Y - \bar{Y})$

Regression coefficient of x on y

$$bxy = \frac{r\sigma x}{\sigma y} = \frac{\Sigma xy}{\Sigma y^2}$$

y on x

$Y - \bar{Y} = byx\,(X - \bar{X})$

Regression Coefficient of y on x

$$byx = \frac{r\sigma y}{\sigma x} = \frac{\Sigma xy}{\Sigma x^2}$$

Deviations from Assumed mean

x on y

$X - \bar{X} = bxy\,(Y - \bar{Y})$

$$bxy = \frac{N\Sigma dx\,dy - \Sigma dx \cdot \Sigma dy}{N\Sigma dy^2 - (\Sigma dy)^2}$$

y on x

$$byx = \frac{N\Sigma dx\,dy - \Sigma dx \cdot \Sigma dy}{N\Sigma dx^2 - (\Sigma dx)^2}$$

Q2

Sol

X	$X-\bar{X}$ (x)	x^2	Y	$Y-\bar{Y}$ (y)	y^2	xy
1	-2	4	6	4	16	-8
5	2	4	1	-1	1	-2
3	0	0	0	-2	4	0
2	-1	1	0	-2	4	2
1	-2	4	1	-1	1	2
2	-1	1	2	0	0	0
7	4	16	1	-1	1	-4
3	0	0	5	3	9	0
					36	-10

$\Sigma X = 24$

$\bar{X} = \frac{\Sigma X}{N} = \frac{24}{8} = 3$

$\bar{Y} = \frac{\Sigma Y}{N} = \frac{16}{8} = 2$

Regression equation X on Y

$X - \bar{X} = bx(Y - \bar{Y})$

$bxy = \frac{\Sigma x}{\Sigma y^2} = \frac{-10}{36} = -0.28$

$(X - 3) = -0.28(Y - 2)$

$X - 3 = -0.28Y + 0.56$

$X = -0.28Y + 0.56 + 3$

$X = 3.56 - 0.28Y$

Regression equation of Y on X

$Y - \bar{Y} = byx(X - \bar{X})$

$byx = \frac{\Sigma xy}{\Sigma x^2} = \frac{-10}{30} = -0.333$

$Y - 2 = -0.333(X - 3)$

$Y - 2 = -0.333X + 0.999$

$Y = -0.333X + 0.999 + 2$

$Y = 2.999 - 0.333X$