Chapter 3 Part 2
REGRESSION ANALYSIS

Meaning:

When it is known that two variables are closely related or correlated, the estimation of one variable, given the value of other variable can be made.

"It is a statistical tool with the help of which we are in a position, to estimate the unknown value of one variable, from known values of another variable is known as Regression".

The dictionary meaning of the term Regression is 'the act of returning back' or stepping back'. Literally regression is opposite to the word 'Progression'.

In statistics regression is used to estimate the probable trend in a related variable when the value of other related variable is given, taking into the account, the past trend.

The term Regression was first used by British Biometrician *SIR FRANCIS GALTON* in 19^{th} century.

Regression Analysis:

It is a technique which measures the average relationship between two or more variables And helps in estimating the value of dependent variable given the value of independent variable.

Dependent Variable:

It is one whose value is influenced by other variables. In other words the variable which is to be estimated/ predicted is called dependent variable.

Independent Variable:

It is one which influences the value of the other related variable. In other words the variables which is used for estimation or prediction is called Independent variable.

Regression Lines/ Estimating Lines:

The line of regression is the straight line which gives the best estimate of one variable for any given value of the other variable. It is a graphic technique to show the functional relationship between two variables.

Ex: - X is dependent variable & Y is independent variable. It shows the average relationship between the two variables X&Y. (& Vice-versa)

In case of two variable X&Y as stated earlier we shall have two regression lines they are: -

a) The regression line of X on Y

b) The regression line of Y on X

Regression Equations:

Regression equations are algebraic expressions of the regression lines, for each of the regression line has regression equations.

a) The regression equation of Y on X

b) The regression equation of X on Y

a) The regression equation of Y on X

$$Y_c = a+bX$$

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

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

b) The regression equation of X on Y

$X_c = a + bY$

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

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

Regression Equations through Correlation Coefficient:

A regression coefficient represents the increment in the value of dependent value for unit change in the value of independent variable. In other words, Regression coefficient represents the rate of change of dependent variable with respect to the rate of change of independent variable.

In two regression equations, **$Y_c = a+bX$ & $X_c = a+bY$** is regarded as regression coefficient. As there are two regression lines, two regression equations, there are two regression coefficient.

a) The regression coefficient of Y on X

$$b_{yx} = \frac{Cov(x,y)}{V(y)} = r\frac{\sigma_y}{\sigma_x}$$

b) The regression coefficient of X on Y

$$b_{xy} = \frac{Cov(x,y)}{V(x)} = r\frac{\sigma_x}{\sigma_y}$$

<u>When deviations are taken from actual mean:</u>

a) The regression coefficient of Y on X under actual mean

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

Thus the regression equation of Y on X can be written as $(Y-\bar{Y}) = b_{yx}(X-\bar{X})$

b) The regression coefficient of X on Y under actual mean

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

Thus the regression equation of X on Y can be written as $(X-\bar{X}) = b_{xy}(Y-\bar{Y})$

<u>When deviations are taken from assumed mean:</u>

a) The regression coefficient of Y on X under assumed mean

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

b) The regression coefficient of X on Y under assumed mean

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

<u>Calculation of Correlation coefficient from Regression Coefficient:</u>

Calculation of r = $\sqrt{bxy \times byx}$

****

PROBLEMS ON REGRESSION

A) Problems on Regression Equation.

1. Find the two regression equations using simultaneous equations method.

X	10	6	5	4	3	2
Y	11	7	6	5	4	3

2. From the following data obtain the two regression equations.

X	3	6	9	10	7
Y	11	8	7	6	8

Hence predict Y if X = 5 and X if Y = 10

B) Problems on Regression equations through Regression coefficients.

I. Actual Mean

1. From the following data obtain the regression equation X on Y and the regression equation Y on X.

X	6	4	8	10	2
Y	9	8	7	5	11

2. Formulate both the regression lines from the following data. Predict Y when X = 50 and X when Y = 25.

X	40	32	38	42	36	46
Y	30	35	40	36	28	35

3. The following data relate to the ages of Husband and Wives.

Age of Husbands (Years)	25	28	30	32	35	36	38	39	42	55
Age of Wives (Years)	20	26	29	30	25	18	26	35	35	46

Find KPCC and most likely age of husband when wife's age is 25 years.

4. Obtain the two regression equations from the following data when the deviations are obtained from mean.

$N=20; \overline{X}=4; \overline{Y}=2; \sum x^2=1680; \sum y^2=320; \sum xy=480.$

II. Assumed Mean

1. The following table shows the age (X) and blood Pressure (Y) of 6 Persons.

X	52	45	36	72	65	47
Y	62	53	51	75	79	60

Obtain two regression equations. Also find the expected blood pressure of person who is 50 years old.

2. Given the bivariate data:

X	1	2	3	5	1	1	3	7
Y	6	0	0	1	1	2	5	1

(a) Fit a regression line Y on X and hence predict Y if X = 5

(b) Fit a regression line X on Y and hence predict X if Y = 2.5

C) Calculation of Correlation coefficient from regression coefficients.

Condition: a) Value of 'r' cannot exceed ±1.

b) Both regression coefficients should have same signs. (Different signs cannot be computed.)

c) If both regression coefficients are negative, 'r' should also be negative. (Though it is positive)

1. Calculate 'r' if b_{xy} = 0.8 and b_{yx} = 0.96.
2. Calculate 'r' if b_{xy} = 0.36 and b_{yx} = - 1.38.
3. Calculate 'r' and two regression coefficients when r = 0.9; $^{\sigma}x$ = 10, $^{\sigma}y$ = 1.5
4. If one of the regression coefficients is 1.5 and 'r' is 0.55, find the value of the other regression coefficient.
5. Find 'r'.

Consolidated Final 14 marks Problems:

1. From the following data,

a) The two regression coefficients
b) The two regression equations
c) The coefficient of correlation between the marks in statistics and mathematics
d) The most likely marks in mathematics when marks in statistics are 30.

Marks in statistics	25	28	35	32	31	36	29	38	34	32
Marks in mathematics	43	46	49	41	36	32	31	30	33	39

2. From the following data you are required to:

a) Form two regression equations
b) Estimate the value of X when Y =10 and the value of Y when X =15
c) Find correlation coefficient through regression coefficients

Particulars	Series X	Series Y
Number of pairs of Observations	5	5
Mean	6	8
Sum of squares of deviation of X and Y from their respective means	40	20
Sum of the product of deviations of X and Y from their respective mean	-26	

3. The data about the sales and advertisement expenditure of a firm are given below.

Particulars	Sales (in Crore Rs.)	Advertisement Expenses (in Crore Rs.)
Mean	40	6
Standard Deviation	10	1.5
Coefficient of Correlation	0.9	

i) Estimate the likely sales for proposed advertisement expenses of Rs.10 Crores.
ii) What should be the advertisement expenses if the firm proposes a sales target of Rs. 60 Crores?

4. The following results of capital employed and profit earned by a firm in 10 successive years are calculated:

Particulars	Mean	Standard Deviation
capital employed (Rs. in '000)	55	28.7
profit earned (Rs. in '000)	13	8.5
Coefficient of Correlation	0.96	

i) Obtain the two regression equations
ii) Estimate the amount of profit to be earned if capital employed is Rs. 50000/-
iii) Estimate the amount of capital to be employed if profit earned is Rs. 20000/-

5. To study the relationship between expenditure on accommodation (X) and expenditure on food and entertainment on food and entertainment (Y) an enquiry into 50 families gave the following results.

$\sum X = 8500$; $\sum Y = 9600$; $\sigma x = 60$; $\sigma y = 20$; 'r' = 0.6.

ASSIGNMENT:

1. You are given

Particulars	X	Y
Mean	18	100
Standard Deviation	14	20
Coefficient of Correlation	0.8	

Find out most probable value of Y if X is 70 and that of X if Y is 90.

2. You are given

Particulars	X	Y
Mean	18	100
Variance	14	20
Coefficient of Correlation	0.8	

Compute the regression line X on Y and Calculate X when Y is 88.

3. The following data relate to marks obtained by 250 student in Accountancy and statistics in B.Com Examination of BCU.

Subject	Mean	Variance
Accountancy	48	16
statistics	55	25
Coefficient of Correlation	0.8	

(1) Find two regression equations.

(2) Estimate the marks obtained by a student in statistics who secured 50 marks in Accountancy, and

(3) Estimate the marks obtained in Accountancy when marks in statistics are 65.

****

3.7 PRACTICALS ON SKILLS DEVELOPMENT ON "BDA"

Chapter-1> Draw a bivariate table using imaginary data.
Chapter-2> For imaginary data of 50 student's marks in 'Business Data Analysis', Compute measures of central tendency.
Chapter-3> For imaginary data of any two variables, calculate 'coefficient of correlation'.
Chapter-4> Collect the sales data of a company for 9 years and estimate the trend values.
Chapter-5> Based on imaginary 5 years data of Production & Sales of a company, Extrapolate the value of Variable for next year.

****