## 1. Introduction

Welcome to the series of E-learning modules on practical problems on regression of two variables. Here we fit a linear regression of two variables. We also find the mean and product moment correlation coefficient from two linear regression lines.

By the end of this session, you will be able to

- Fit a regression line of Dependent variable on Independent variable
- Find the means of the variables using regression lines
- Find coefficient of product moment correlation using regression coefficient.
- Estimate the value of dependent variable, when independent variable is given

In a bivariate data, if X is independent variable and Y is dependent variable, the equation of regression line, Y on X is given by

Y minus y bar is equal to b y, x into x minus x bar

Where, b y, x is equal to r into sigma y divided by sigma x.

r is coefficient of correlation between X and Y, sigma x is standard deviation of variable X, sigma Y is the standard deviation of Y

Similarly, if X is dependent variable and Y is independent variable, then equation for regression line X on Y is given by,

X minus x bar is equal b x y into y minus y bar

Where, b x y is equal to r into sigma x divided by sigma y.

r is coefficient of correlation between X and Y, sigma x is standard deviation of variable X, sigma Y is the standard deviation of Y

suppose we have given raw data, then b x y is give by,

b x y is equal to n into summation x into y minus summation x into summation y divided by n into summation y square minus summation y the whole square.

Similarly, b y x is equal to n into summation x into y minus summation x into summation y divided by n into summation x square minus summation x the whole square.

In the case of tabulated data,

b x y is equal to N into double summation f into x into y minus summation f into x into summation f into y divided by N into summation f into y square minus summation f into y the whole square.

Similarly, b y x is equal to N into double summation f into x into y minus summation f into x into summation f into y divided by N into summation f into x square minus summation f into x the whole square.

Given the regression equations, X on Y and Y on X, we can find the means x bar and y bar by solving the two simultaneous equations.

We can identify the equations X on Y and Y on X, and can find the regression coefficients and in turn we can find correlation coefficient between the two variables X and Y.

To obtain the graph of the regression equation,

- We find two points on the regression line.
- These points are plotted on a graph sheet

• The points are joined by straight line The resulting line is regression line of x on y

### 2. Exercises

### Exercise one

Find the most likely price in Mumbai corresponding to price of seventy rupees at Kolkata from the following

### Figure 1

	Kolkata	Mumbai
Average Price	65	67
Standard Deviation	2.5	3.5

Average Price of Kolkata and Mumbai are 65 and 67 respectively. Also the respective Standard Deviations are 2 point 5 and 3 point 5

Correlation coefficient between the prices of commodities in the two cities is zero point 8

Solution

Let the prices (in Rupees) in Kolkata and Mumbai be denoted by X and Y respectively. Then we have given, x bar is equal to 65, y bar is equal to 67, sigma x is equal to 2 point 5, sigma y is equal to 3 point 5, r is equal to zero point 8.

Line of regression Y on X is

Y minus y bar is equal to r into sigma y divided by sigma x into x minus x bar

That is, y minus 67 is equal to zero point 8 into 3 point 5 divided by 2 point 5 into x minus 65.

Hence Y is equal to 1 point 1, 2 into X minus 5 point 8

When X is equal to 70, Y is equal to 1 point 1, 2 into 70 minus 5 point 8 Is equal to 72 point 6

Hence the most likely price in Mum

Hence the most likely price in Mumbai corresponding to the price of seventy rupees at Kolkata is 72 point six zero rupees.

### Exercise two

A survey of children revealed the following information regarding IQ of child and age of the mother at the time of giving birth to child.

	IQ of Child	Age of Mother
Mean	98	28 years
Standard Deviation	2	4 years

### Figure 2

Mean IQ and age of the mother is 98 and 28 years respectively. And respective Standard deviations are 2 and 4 years.

Coefficient of correlation r is equal to minus zero point two four

Estimate IQ of a child whose mother was aged 47 at the time of giving birth to the child.

Solution

Let x denote IQ of a child and y denote the age of the mother at the time of giving birth to the child.

Given x bar is equal to 98, y bar is equal to 28, r is equal to minus zero point two four, Standard deviation of X is equal to 2; Standard deviation of y is equal to 4.

We assume that IQ depends on age of the mother. And so for the estimation of x given y we write down regression equation of x on y.

B x y is equal to r into sigma x divided by sigma y is equal to minus zero point 2two three into 2 divided by 4 is equal to minus zero point one two.

The regression equation of x on y is

X minus x bar is equal to b x y into y minus y bar

On substitution,

X minus 98 is equal to minus zero point one two into y minus 28

Implies, x is equal to minus zero point one two into y plus one hundred and one point three six The estimate of the I Q of a child when the mother aged 47 while giving birth to the child, is given by

X is equal to zero point one two into 47 plus one hundred and one point three six is equal to 95 point five six

To plot the regression line we find 2 points, that is when y is equal to 20, x is equal to 98 point 9, 6 and y is equal to 40, x is equal to 96 point five six.

The points are plotted and they are joined by a straight line to get regression line of x on y as follows.

The line in the graph shows the regression line of X on Y.

### Figure 3



### 3. Exercises Contd.

#### Exercise 3

In a bivariate data on x and y the means are respectively 15 and 27 and the variances are respectively 25 and 9. The coefficient of correlation is minus zero point 3. Then what would be the expected value of y when x is equal to 8.

### Figure 4

	Mean	Variance
x	15	25
У	27	9

#### Solution

Here we have given x bar is equal to 15, y bar is equal to 27,

Sigma x is equal to square root of 25 is equal to 5, sigma y is equal to square root of 9 is equal to 3 and r is equal to minus zero point 3

B y x is equal to r into sigma y divided by sigma x is equal to minus zero point 3 into 3 divided by 5 is equal to minus zero point one eight.

Therefore y minus y bar is equal b y x into x minus x bar By substituting, we get y minus 27 is equal minus zero point one eight into x minus 15 implies, y is equal to minus zero point one eight into x plus 29 point 7 when x is equal to 8, the estimate of y is given by y is equal to minus zero point one eight into 8 plus 29 point 7 is equal to 28 point two six.

### Exercise 4

In a bivariate data, summation x is equal to 20, summation y is equal to four hundred, summation x square is equal to one hundred and ninety six, summation y square is equal to 46 thousand and five hundred, summation x into y is equal to eight hundred and fifty and n is equal to ten. Estimate the value of y corresponding to the value of x is equal to five.

### Solution

Since we need to find the value of y corresponding to given value of x, we should find the regression equation of y on x.

First we find the means and regression coefficients as follows. X bar is equal to summation x divided by n is equal to 30 by 10 is equal to 3

Y bar is equal to summation y divided by n is equal to four hundred by 10 is equal to forty

B y x is equal to n into summation x into y minus summation x into summation y divided by n into summation x square minus summation x the whole square

is equal to 10 into 850 minus 30 into 40 divided by 10 into 196 minus 30 square is equal to minus 3 point 3

the regression equation of y on x is,

y minus y bar is equal to b y x into x minus x bar on substitution, we get, y minus 40 is equal to minus 3 point 3 into x minus 3 implies y is equal to minus 3 point 3 into x plus 49 point 9 The estimate of y when x is equal to 5 is given by Y is equal to minus 3 point 3 into 5 plus 49 point 9 Is equal to 33 point 4.

## 4. Exercises on Regression Equations Contd.

### Exercise 5

The following are heights of 8 persons and one each of their sons. From the data, estimate the height of a person whose father is 150 cms tall.

The table gives the heights of 8 fathers and corresponding heights of theirs son.

### Figure 5

Height of Father	164	176	178	184	175	167	173	180
Height of Son	168	174	175	181	173	166	173	179

### Solution

Let x and y respectively denote the heights of fathers and the sons. Then the value of y corresponding to x is equal to 150 has to be estimated. For this, regression of y on x should be found and the estimation should be made.

Let us find the following table which gives required totals of the different columns.

### Figure 6

Height of Father (x)	Height of Son (y)	x <sup>2</sup>	ху	
164	168	26896	27552	
176	174	30976	30624	
178	175	31684	31150	
184	181	33856	33304	
175	173	30625	30275	
167	166	27889	27722	
173	173	29929	29929	
180	179	32400	32220	
1397	1389	244255	242776	

The first and second columns are written as it is in the given problem. The third column is obtained by squaring the numbers in the first column. That is 164 is equal to 26 thousand 8hundred and 96, 176 square is equal to 30 thousand 9 hundred and 76 etc. The last column is obtained by multiplying the numbers in first and second column.

That is 164 into 168 is equal to 27 thousand 5 hundred and 52

176 into 174 is equal 30 thousand 6 hundred and 24 etc.

The last row, which is written in bold gives the total of each column.

Now let us find means and regression coefficients.

X bar is equal to summation x divided by n is equal to one thousand 3hundred and 97 divided by 8 is equal to 174 point 6, 3 and

Y bar is equal to summation y divided by n is equal to one thousand 3 hundred and 89 divided by 8 is equal to 173 point 6, 3

Regression coefficient b y x is equal to n into summation x into y minus summation x into summation y divided by n into summation x square minus summation x the whole square is equal to 8 into 2 lakh 42 thousand 776 minus 1 thousand 397 into 1 thousand 389 divided by 8 into 2 lakh 44 thousand 2 hundred and 55 minus 1 thousand 3 hundred and 97 the whole square is equal to zero point 7, 3 zero 2

Thus regression of y on x is,

Y minus y bar is equal to b y x into x minus x bar

On substitution, we get, y minus 173 point 6, 3 is equal to zero point 7, 3 zero 2 into x minus 174 point 6, 3

Implies, y is equal to zero point 7, 3 zero 2 into x plus 46 point 1, 2.

To estimate the value of y when x is equal to 150 substitute the value for x as 150 in the above equation.

Thus, the estimate of son's height is,

Y is equal to zero point 7, 3, zero 2 into 150 plus 46 point 1, 2

Is equal to 155 point 6, 5 centimetres.

### Exercise 6

The following table gives agricultural production index and the index of wholesale price in 5 years.

### Figure 7

Agricultural production index	104	110	112	114	120
Wholesale price index	106	116	140	175	173

Find the regression equation of wholesale price index on the agricultural production index and estimate the wholesale price index when the agricultural production index is 125

### Solution

Let x denote the agricultural production index and y denote the wholesale price index. Here we need to estimate the wholesale price index when the agricultural production index is 125. Hence we should find the regression equation of y on x.

Let us construct the following table which gives required totals of the different columns.

### Figure 8

Agricultural Production	Wholesale Price	X <sup>2</sup>	ху
Index	Index		
104	106	10816	11024
110	116	12100	12760
112	140	12544	15680
114	175	12996	19950
120	173	14400	20760
560	710	62856	80174

The first and second columns are written as it is in the question, which denote x and y respectively.

The third column is obtained by squaring the numbers in the 1<sup>st</sup> column. That is 104 square is equal to 10 thousand 816, 110 square is equal to 12 thousand 100 etc. The last column is obtained by multiplying the numbers in first and second columns. That is 104 into 106 is equal to 11 thousand 24, 110 into 116 is equal to 12 thousand 760 etc. The last row, which is denoted in bold numbers give the totals of each column.

V han is a smaller and stick a level is a smaller 500 divided by 5 is a smaller 440

X bar is equal to summation x by n is equal to 560 divided by 5 is equal to 112 Y bar is equal to summation y by n is equal to 710 divided by 5 is equal to 142

B y x is equal to n into summation x into y minus summation x into summation y divided by n into summation x square minus summation x the whole square.

Is equal to 5 into 80 thousand 174 minus 560 into 710 divided by6 5 into 62 thousand 856 minus 560 the whole square is equal to 4 point 8 zero 9

The regression equation of y on x is,

Y minus y bar is equal to b y x into x minus x bar

On substitution, we get

Y minus 142 is equal to 4 point 8 zero 9 into x minus 112

On simplification we get,

Y is equal to 4 point 8 zero 9 into x minus 396 point 6 zero 8

The estimate of wholesale price index(y) when agricultural price index x is equal to 125 is given by,

Y is equal to 4 point 8 zero 9 into (125) minus 396 point 6 zero 8 Is equal to 204 point 5, 1, 7

# 5. Exercises on Coefficient of Correlation

### Exercise 7

In a bivariate data, the regression coefficients are minus 7 point 3 and minus zero point one one. Find the coefficient of correlation

Solution

The coefficient of correlation is numerically equal to the geometric mean of the regression coefficients.

Therefore the coefficient of correlation is given by, r is equal to plus or minus b x y into b y x is equal to plus or minus square root of minus 7 point three into minus zero point one one is equal to minus zero point eight nine six one

Since the regression coefficients are negative, the coefficient of correlation is also negative. Thus coefficient of correlation is r is equal to minus zero point eight nine six one

### Exercise 8

In a bivariate data, the two regression equations are 3x plus 4y is equal to one and 3x plus y is equal to 4. Find the means and coefficient of correlation.

Solution

Since the regression line intersect at x bar, y bar the values x and y are obtained by solving the regression equations for x and y.

The equations are

3x plus 4y is equal to 1 name it as equation number (1)

3x plus y is equal to 4 name it as equation number (2)

On subtracting we get,

Three y is equal to minus three or y is equal to minus one

Putting y is equal to minus 1 in equation 1 we get,

three x minus four is equal to one or three x is equal to five or x is equal to five by three

Thus x bar is equal to five by three and y bar is equal to minus one.

Now let us consider the two equations as y on x and x on y.

Let regression equation of y on x be three x plus four y is equal to one and regression equation of x on y be three x plus y is equal to four

5 1 7 1

The equations are re-written as

Y is equal to (minus three by four) into x plus one by four

And x is equal to (minus one by three) into y plus (four by three)

From above equations,

b y x is equal to (minus three by for) and b x y is equal to (minus one by three)

We know that r is geometric mean of regression coefficients. i.e., r is equal to plus or minus square root of b x y into b y x is equal to plus or minus square root of minus 1 by 3 into minus 3 by 4 is equal to minus half. Observe that r is negative, since both the regression coefficients are negative.

Note: When we consider arbitrarily the given equations as X on Y and Y on X and we get r greater than 1 then we interchange the equations. that is, the equation X on Y should be considered as Y on X and the equation Y on X should be considered as X on Y.

Here's a summary of our learning in this session where we have understood how to:

- Fit a regression line of dependent variable on independent variable
- Find the means of the variables using regression lines
- Find coefficient of product moment correlation using regression coefficient
- Estimate the value of dependent variable, when independent variable is given