Problem 3: From the following results, obtain the two regression equations and estimate the yield of crops when the rainfall is 29 cms and the rainfall when the yield is 600 kg.

	Y (yield in kgs)	X (rainfall in cms)
Mean	508.4	26.7
S.D.	36.8	4.6

Coefficient of correlation between yield and rainfall is 0.52.

Solution 3: We have $\overline{x} = 26.7, \overline{y} = 508.4, \sigma_x = 4.6, \sigma_y = 36.8$ and $\rho = 0.52$. Now,

$$b_{yx} = \rho\frac{\sigma_y}{\sigma_x} = 4.16 \quad \text{and} \quad b_{xy} = \rho\frac{\sigma_x}{\sigma_y} = 0.065.$$

The required regression equations are

$$\begin{aligned} y &= 4.16x + 397.328 \\ \text{and } x &= 0.065y - 6.346. \end{aligned}$$

When $x = 29$ cm, we have $y = (4.16 \times 29) + 397.328 = 517.968$ kg.

When $y = 600$ kg, we have $x = (0.065 \times 600) - 6.346 = 32.654$ cm.

i.e., when the rainfall is 29 cms, the yield of the crop is 517.968 kg, and when the yield is 600 kg, the rainfall is 32.654 cms.

Multiple linear Regressions

If the number of independent variables in a regression model is more than one, then the model is called as multiple regression. In fact, many of the real-world applications demand the use of multiple regression models.

Regression Model with Two independent variables using Normal equations:

Suppose the number of independent variables is two, then

$$Y = b_o + b_1 X_1 + b_2 X_2$$

Normal equations are

$$\Sigma Y = n b_o + b_1 \Sigma X_1 + b_2 \Sigma X_2$$

$$\Sigma Y X_1 = b_o \Sigma X_1 + b_1 \Sigma X_1^{\,2} + b_2 \Sigma X_1 X_2$$

$$\Sigma Y X_2 = b_o \Sigma X_2 + b_1 \Sigma X_1 X_2 + b_2 \Sigma X_2^{\,2}$$

Solving for b_0, b_1 and b_2, we get the Multiple regression equation:

$$\boxed{Y = \bar{y} + r_{YX_1.X_2}(X_1 - \overline{x_1}) + r_{YX_2.X_1}(X_2 - \overline{x_2})}$$

Problem 1: The annual sales revenue(in crores of rupees) of a product as a function of sales force(number of salesmen) and annual advertising expenditure(in lakhs of rupees) for the past 10 year are summarized in the following table.

Annual sales revenue Y	**20**	**23**	**25**	**27**	**21**	**29**	**22**	**24**	**27**	**35**
Sales force X1	**8**	**13**	**8**	**18**	**23**	**16**	**10**	**12**	**14**	**20**
Annual advertising expenditures X2	**28**	**23**	**38**	**16**	**20**	**28**	**23**	**30**	**26**	**32**

Let the regression model be $Y = b_o + b_1X_1 + b_2X_2$

Y	X1	X2	$X1^2$	$X2^2$	X1X2	YX1	YX2
20	8	28	64	784	224	160	560
23	13	23	169	529	229	299	529
25	8	38	64	1444	304	200	950
27	18	16	324	256	288	486	432
21	23	20	529	400	460	483	420
29	16	28	256	784	448	464	812
22	10	23	100	529	230	220	506
24	12	30	144	900	360	288	720
27	14	26	196	676	364	378	702
35	20	32	400	1024	640	700	1120
$\sum y = 253$	$\sum X1 = 142$	$\sum X2 = 264$	$\sum X1^2 = 2246$	$\sum X2^2 = 7326$	$\sum X1X2 = 3617$	$\sum YX1 = 142$	$\sum YX2 = 142$

Substituting the required values in the normal equations, we get the following simultaneous equations

$$253 = 10b_o + 142b_1 + 264b_2$$

$$3678 = 142b_o + 2246b_1 + 3617b_2$$

$$6751 = 264b_o + 3617b_1 + 7326b_2$$

The solution to the above set of simultaneous equation is

$b_o = 5.1483, \quad b_1 = 0.6190 \quad and \quad b_2 = 0.4304$

Therefore, the regression model is $Y = 5.1483 + 0.6190X_1 + 0.4304X_2$