Chapter 15

Life can only be understood backward, but it must be lived forward.

—Niels Bohr

Other than food and sex, nothing is quite as universally interesting as the size of our pay cheques.

—N. L. Preston and E. R. Fiedler

Partial Correlation, Multiple Correlation and Regression Analysis

LEARNING OBJECTIVES

After studying this chapter, you should be able to

- describe the relationship between two variables when influence of one or more other variables is held constant or involved.
- establish a regression equation for estimating value of a dependent variable given the values of two or more independent variables.
- determine the extent of random error involved in the estimation of dependent variable value.
- measure the coefficient of determination to understand the proportion of variation in the dependent variable which is explained by the independent variables.

15.1 INTRODUCTION

The techniques of linear regression analysis for estimating the value of a dependent variable can be extended by involving more than one independent variable (predictors) in a regression equation. This equation can be used to explain with more accuracy the influence of independent variables on the predicted value of the dependent variable.

A linear regression equation with more than one independent variable is referred to as *multiple linear regression models*. Although a multiple linear regression analysis in many ways is an extension of simple linear regression analysis but requires more computational time while estimating the value of the dependent variable. Thus, problems involving only two or three independent variables are discussed in this chapter in order to demonstrate calculations involved in the multiple regression analysis.

A *multiple correlation analysis* is helpful to measure the degree of association between a dependent (response) variable, y, and two or more independent variables (predictors) x_1, x_2, ... taken together as a group. In a multiple correlation analysis, it is also possible to measure the degree of association between a dependent variable and any one of the independent variables included in the analysis, while the effect of the other independent variables included in the analysis is held constant. This measure of association is called *partial correlation coefficient*. It differs from a simple correlation coefficient in the manner that in simple correlation analysis the effect of all other variables is ignored rather than being statistically controlled as in partial correlation analysis.

The main advantage of a multiple regression analysis is that it accommodates more than one independent variable to estimate the value of the dependent variable. The data on the values of independent variables enable decision makers to determine the statistical error associated with the estimated value of the dependent variable, and hence the relationship between a dependent variable and two or more independent variables can be described with more accuracy.

Illustrations The following examples illustrate a multiple regression model (or equation):

1. Suppose a farmer who wishes to relate the yield (y) of wheat crop, a dependent variable, to the amount of fertilizer used, an independent variable, he can certainly find a simple regression equation that relates these two variables. However, true prediction of yield of crop is possible by including in the regression equation a few more variables such as quality of seed, amount of water given to the crop and so on.
2. Suppose if the management of an organization is wishes to understand the expected variance in the performance (y), a dependent variable, to be explained by four independent variables, such as pay, task difficulty, supervisory support and organizational culture. These four independent variables are, of course, correlated to the dependent variable in varying degrees but they might also be correlated among themselves, e.g. task difficulty is likely to be related to supervisory support, pay is likely to be related to task difficulty. These three variables together are likely to influence the organizational culture.

15.2 ASSUMPTIONS IN A MULTIPLE LINEAR REGRESSION

A multiple regression equation is used in the same way as that a simple linear regression equation for understanding the relationship among variables of interest and for estimating the average value of the dependent variable y given a set of values of independent variables. The summary of assumptions for a multiple linear regression is as follows:

1. The linear regression equation involving a dependent variable y and a set of k independent variables $x_1, x_2, \ldots, x_k$ is given by

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + e \tag{15-1}$$

where y = value of dependent variable to be estimated

β_0 = a constant which represents the value of y when value of all independent variables is zero.

$\beta_1, \ldots, \beta_k$ = parameters or *regression coefficients* associated with each of the x_k independent variables.

x_k = value of kth independent variable.

e = random error associated with the sampling process. It represents the unpredictable variation in y values from the population regression model

The term *linear* is used because equation (15-1) is a linear function of independent variables $x_1, x_2, \ldots, x_k$.

2. The variance and standard deviation of the dependent variable are equal for each combination of values of independent variables.
3. The random error (e) associated with the dependent variable, y, for various combinations of values of independent variables is statistically independent of each other and normally distributed.
4. The error of variance is same for all values of independent variables. Thus, the range of deviations of the y-values from the regression line is same regardless of the values of the independent variables.

   The magnitude of error of variance (also called residual variance) measures the closeness of observed values of the dependent variable to the regression line (line of 'best fit'). The smaller the value, the better the predicted value of the dependent variable.
5. The random error, e, is a random variable with zero expected (or mean) value. Consequently, the expected or mean value of the dependent variable is denoted by

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$$

The parameter $\beta_j (j = 0, 1, 2, ..., k)$ is also called *partial regression coefficient* because it measures the expected change in response variable, y per unit change in the independent variable, x_j when all remaining independent variables, $x_i (i \neq j)$ are held constant.

15.3 ESTIMATING PARAMETERS OF A MULTIPLE REGRESSION MODEL

A multiple regression analysis requires that a sample drawn from the population of interest to calculate values of the unknown parameters $\beta_0, \beta_1, ..., \beta_k$ by fitting the regression line based on the sample data. To fit the general linear multiple regression line using method of least squares, we choose the estimated regression line (or model)

$$E(y) \text{ or } \hat{y} = b_0 + b_1 x_1 + b_2 x_2 + ... + b_k xk \quad (15\text{-}2)$$

that minimizes the ***sum of squares errors (SSE)*** $= y_i - \hat{y}_i$, where y_i and $\hat{y}_i$ $(i = 1, 2..., k)$ represent the observed and estimated (or predicted) value of the dependent variable for the ith observation. The terms $b_j (j = 0, 1, 2, ..., k)$ are the least-squares estimates of population regression parameter βj.

15.3.1 Estimation: The Method of Least Squares

The method of least squares discussed in Chapter 14 to compute the value of regression coefficients (or parameters) a and b and estimating the value of the dependent variable, y in a linear regression line can be extended for estimating the unknown parameters $\beta_0, \beta_1, \beta_2, ..., \beta_k$. based on sample data. The least-squares estimators of these parameters are denoted by $b_0, b_1, b_2, ..., b_k$, respectively. Given these values, the least-squares *multiple regression equation* can be written as

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + ... + b_k x_k \quad (15\text{-}3)$$

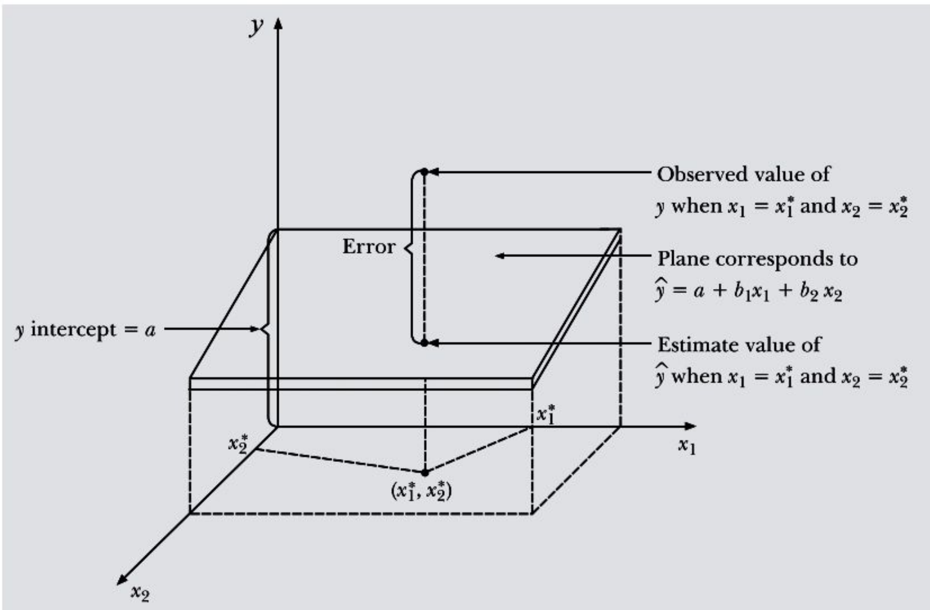
where $\hat{y}$ is estimated value of dependent variable y, b_0 is y-intercept, $x_1, x_2, ..., x_k$ is independent variables and b_j is the slope associated with variable $x_j (j = 1, 2, ..., k)$.

To understand a multiple regression line, consider the following regression equation involving two independent variables x_1 and x_2 and a dependent variable y:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2$$

This equation describes a plane in a 3-dimensional space of y, x_1 and x_2 as shown in Fig. 15.1.

Figure 15.1
Graph of a Multiple Regression Equation with Two Independent Variables



15.3.2 Partial Regression Coefficients

The partial regression coefficient, $\beta_j (j = 1, 2, ..., k)$, represents the expected change in the dependent (response) variable y with respect to per unit change in independent variable, x_j, when all other independent variables $x_i (i \neq j)$ are held constant. The partial regression coefficients occur because more than one independent variable is present in the regression line.

Least-Squares Normal Equations

To demonstrate the calculation of partial regression coefficients, consider a regression equation involving two independent variables x_2 and x_3 and a dependent variable x_1:

$$\hat{x}_1 = b_0 + b_1 x_2 + b_2 x_3 = a + b_{12.3} x_2 + b_{13.2} x_3 \quad (15\text{-}4)$$

where $\hat{x}_1$ is estimated value of dependent variable, b_0 is a regression constant representing intercept on y-axis; its value is zero when the regression equation passes through the origin and $b_{12.3}$, $b_{13.2}$ is partial regression coefficients; $b_{12.3}$ corresponds to change in x_1 for each unit change in x_2 while x_3 is held constant; $b_{13.2}$ represents the change in x_1 for each unit change in x_3 while x_2 is held constant.

The values of the constants, b_0(or $b_{1.23}$), $b_{12.3}$ and $b_{13.2}$ are determined in accordance with the least-squares criterion, i.e. minimizing the sum of the squares of the residuals. Thus,

$$\text{Min } z = \Sigma(x_1 - \hat{x}_1)^2 = \Sigma(x_1 - b_0 - b_{12.3}x_2 - b_{13.2}x_3)^2$$

To minimize this sum, we use the concept of maxima and minima, where derivatives of z with respect to these constants are equated to zero. Consequently, we get the following three normal equations:

Estimated Multiple Regression Equation: The estimate of the multiple regression equation based on sample data and the least-squares method.

$$\begin{aligned} \Sigma x_1 &= n\, b_0 + b_{12.3}\, \Sigma x_1 + b_{13.2}\, \Sigma x_3 \\ \Sigma x_1 x_2 &= b_0 \Sigma x_2 + b_{12.3} \Sigma x_{22} + b_{13.2} \Sigma x_3\, x_2 \\ \Sigma x_1 x_3 &= b_0 \Sigma x_3 + b_{12.3} \Sigma x_2 x_3 + b_{13.2}\, \Sigma x_{32} \end{aligned} \quad (15\text{-}5)$$

The value of constants can be calculated by solving the system of simultaneous equations (15-5).

Short-cut Method

For solving above stated normal equations for constants, e.g. $b_0, b_{12.3}$ and $b_{13.2}$, take deviations of the values of the variables from their actual mean values. Let $X_1 = x_1 - \bar{x}_1$, $X_2 = x_2 - \bar{x}_2$ and $X_3 = x_3 - \bar{x}_3$ be the deviations from the actual mean values of variables x_1, x_2 and x_3, respectively. Since the sum of deviations of the values of a variable from its actual mean is zero, therefore

$$\Sigma X_1 = \Sigma(x_1 - \bar{x}_1) = 0,\ \Sigma X_2 = (x_2 - \bar{x}_2) = 0, \text{ and } \Sigma X_3 = (x_3 - \bar{x}_3) = 0$$

Summing the variables and dividing by n in the regression equation of x_1 on x_2 and x_3, we have

$$\bar{x}_1 = b_{1.23} + b_{12.3}\,\bar{x}_2 + b_{13.2}\,\bar{x}_3 \quad (15\text{-}6)$$

Subtracting equation (15-6) from equation (15-4), we have

$$\begin{aligned} x_1 - \bar{x}_1 &= b_{12.3}(x_2 - \bar{x}_2) + b_{13.2}(x_3 - \bar{x}_3) \\ X_1 &= b_{12.3} X_2 + b_{13.2}\, X_3 \end{aligned}$$

The second and third equations of (15-5) can be rewritten as

$$\begin{aligned} \Sigma\, X_1\, X_2 &= b_{12.3} + b_{13.2}\, \Sigma X_2 X_3 \\ \Sigma\, X_1\, X_3 &= b_{12.3}\, \Sigma X_2 X_3 + b_{13.2} \end{aligned} \quad (15\text{-}7)$$

When simultaneous equations (15-7) are solved for $b_{12.3}$ and $b_{13.2}$, we have

$$b_{12.3} = \frac{(\Sigma X_1X_2)(\Sigma X_3^2) - (\Sigma X_1X_3)(\Sigma X_2X_3)}{\Sigma X_2^2 \times \Sigma X_3^2 - (\Sigma X_2X_3)^2} \quad (15\text{-}8)$$

and

$$b_{13.2} = \frac{(\Sigma X_1X_3)(\Sigma X_2^2) - (\Sigma X_1X_2)(\Sigma X_2X_3)}{(\Sigma X_2^2)(\Sigma X_3^2) - (\Sigma X_1X_3)^2}$$

15.3.3 Relationship Between Partial Regression Coefficient and Correlation Coefficient

Let r_{12} is correlation coefficient between variable x_1 and x_2 and r_{13} = correlation coefficient between variable x_1 and x_3.

These correlation coefficients are known as *zero order correlation coefficients*. The variance of sample data on variables x_1, x_2, and x_3 is given by

$$s_1^2 = \frac{\Sigma (x_1 - \bar{x}_1)^2}{n} = \frac{\Sigma X_1^2}{n}$$

$$s_2^2 = \frac{\Sigma (x_2 - \bar{x}_2)^2}{n} = \frac{\Sigma X_2^2}{n}$$

$$s_3^2 = \frac{\Sigma (x_3 - \bar{x}_3)^2}{n} = \frac{\Sigma X_3^2}{n}$$

Since

$$r_{12} = \frac{\text{Cov.}(x_1, x_2)}{\sigma_{x_1}\sigma_{x_2}} = \frac{\Sigma (x_1 - \bar{x}_1)(x_2 - \bar{x}_2)}{\sqrt{\Sigma (x_1 - \bar{x}_1)^2}\sqrt{\Sigma (x_2 - \bar{x}_2)^2}} = \frac{\Sigma X_1X_2}{\sqrt{\Sigma X_1^2 \, \Sigma X_2^2}}$$

$$= \frac{\Sigma X_1X_2}{\sqrt{n\, s_1^2}\sqrt{n\, s_2^2}} = \frac{\Sigma X_1X_2}{n\, s_1\, s_2}$$

or $\Sigma X_1X_2 = n\, r_{12}\, s_1\, s_2$

Similarly $\Sigma X_1X_3 = n\, r_{13}\, s_1 s_3$ and $\Sigma X_2X_3 = n\, r_{23}\, s_2\, s_3$

Substituting these values in equation (15-7) and on simplification, we get

$$b_{12.3}\, s_2 + b_{13.2}\, s_3\, r_{23} = s_1\, r_{12}$$
$$b_{13.2}\, s_3 + b_{12.3}\, s_2\, r_{23} = s_1\, r_{13} \quad (15\text{-}9)$$

Solving equations (15-9) for $b_{12.3}$ and $b_{13.2}$, we have

$$b_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{1 - r_{23}^2}\left(\frac{s_1}{s_2}\right)$$

and

$$b_{13.2} = \frac{r_{13} - r_{12}\, r_{23}}{1 - r_{23}^2}\left(\frac{s_1}{s_3}\right) \quad (15\text{-}10)$$

Thus, the regression equation of x_1 on x_2 and x_3 can be written as

$$X_1 = b_{12.3}\, X_2 + b_{13.2}\, X_2$$

$$x_1 - \bar{x}_1 = \left(\frac{r_{12} - r_{13}\, r_{23}}{1 - r_{23}^2}\right)\left(\frac{s_1}{s_2}\right)(x_2 - \bar{x}_2) + \left(\frac{r_{13} - r_{12}\, r_{23}}{1 - r_{23}^2}\right)\left(\frac{s_1}{s_3}\right)(x_3 - \bar{x}_3)$$

Example 15.1: A sample survey of 5 families was taken and figures were obtained with respect to their annual savings x_1(₹ in 100's), annual income x_2(₹ in 1000's) and family size x_3. The data is summarized in the table below:

Family	*Annual Savings* (x_1)	*Annual Income* (x_2)	*Family Size* (x_3)
1	10	16	3
2	5	13	6
3	10	21	4
4	4	10	5
5	8	13	3

(a) Find the least-squares regression equations of x_1 on x_2 and x_3.
(b) Estimate the annual savings of a family whose size is 4 and annual income is ₹16,000.

Solution: (a) The calculations for three variables regression problem are shown in Table 15.1.

Table 15.1: Calculations for Regression Equation

Family	*Savings* x_1	*Income* x_2	*Size* x_3	x_1^2	x_2^2	x_3^2	$x_1 x_2$	$x_1 x_3$	$x_2 x_3$
1	10	16	3	100	256	9	160	30	48
2	5	13	6	25	169	36	65	30	78
3	10	21	4	100	441	16	210	40	84
4	4	10	5	16	100	25	40	20	50
5	8	13	3	64	169	9	104	24	39
$n = 5$	37	73	21	305	1135	95	579	144	299

Substituting values in the normal equations for the regression equation of x_1 on x_2 and x_3, we get

(i) $\Sigma x_1 = n\, b_0 + b_1\Sigma x_2 + b_2\Sigma x_3$ or $37 = 5\, b_0 + 73b_1 + 21b_2$

(ii) $\Sigma x_1x_2 = b_0\Sigma x_2 + b_1\Sigma x_{22} + b_2\Sigma x_2x_3$ or $579 = 73\, b_0 + 1135b_1 + 299b_2$

(iii) $\Sigma x_1x_3 = b_0\Sigma x_3 + b1\Sigma x_2x_3 + b_2\Sigma x_{32}$ or $144 = 21\, b_0 + 299b_1 + 95b_2$

Multiply equation (i) by 73 and equation (ii) by 5 to eliminate b_0, we get

(iv) $194 = 346b_1 - 38b_2$

Multiply equation (i) by 21 and equation (iii) by 5 to eliminate b_0, we get

(v) $57 = 38b_1 - 34b_2$

Multiplying equation (iv) by 34 and equation (v) by 38 to eliminate b_2, we get

$5190 = 10320b_1$ or $b_1 = 0.502$

Substituting the value of b_1 in equations (i) and (ii), we get

(vi) $5b_0 + 21b_2 = 0.354$

(vii) $21b_0 + 95b_2 = -6.098$

Multiplying equation (vi) by 21 and equation (vii) by 5 to eliminate b_0, we get

$$-34b_2 = 37.924 \text{ or } b_2 = -1.115$$

Substituting the value of b_1 and b_2 in equation (i), we get $b_0 = 4.753$. (a) The least-squares regression equation of x_1 on x_2 and x_3 is given by

$$x_1 = b_0 + b_1x_2 + b_2\, x_3$$
$$= 4.753 + 0.502x_2 - 1.115\, x_3$$

(b) The estimated value of annual savings (x_1) is obtained by substituting annual income x_2 = ₹1600 and family size $x_3 = 4$ as 4.753 + 0.502(1600) – 1.115(4) = ₹803.493

Example 15.2: An instructor of mathematics wishes to determine the relationship of grades in the final examination to grades on two quizzes given during the semester. Let x_1, x_2 and x_3 be the grades of a student on the first quiz, second quiz and final examination, respectively. The instructor made the following computations for a total of 120 students:

$$\bar{x}_1 = 6.80 \qquad \bar{x}_2 = 0.70 \qquad \bar{x}_3 = 74.00$$
$$s_1 = 1.00 \qquad s_2 = 0.80 \qquad s_3 = 9.00$$
$$r_{12} = 0.60 \qquad r_{13} = 0.70 \qquad r_{23} = 0.65$$

(a) Find the least-squares regression equation of x_3 on x_1 and x_2.
(b) Estimate the final grades of two students who scored respectively 9 and 7 and 4 and 8 marks in the two quizzes. [*HP Univ., MBA, 2005*]

Solution: (a) The regression equation of x_3 on x_2 and x_1 can be written as

$$(x_3 - \bar{x}_3) = \left(\frac{r_{23} - r_{13}r_{12}}{1 - r_{12}^2}\right)\left(\frac{s_3}{s_2}\right)(x_2 - \bar{x}_2) + \left(\frac{r_{13} - r_{23}r_{12}}{1 - r_{12}^2}\right)\left(\frac{s_3}{s_1}\right)(x_1 - \bar{x}_1)$$

Substituting the given values, we have

$$(x_3 - 74) = \left(\frac{0.65 - 0.7 \times 0.6}{1 - (0.6)2}\right)\left(\frac{9}{0.8}\right)(x_2 - 7) + \left(\frac{0.7 - 0.65 \times 0.6}{1 - (0.6)^2}\right)\left(\frac{9}{1}\right)(x_1 - 6.8)$$

$$(x_3 - 74) = \left(\frac{0.65 - 0.42}{0.64}\right)\left(\frac{9}{0.8}\right)(x_2 - 7) + \left(\frac{0.7 - 0.39}{0.64}\right)(9)(x_1 - 6.8)$$

$$(x_3 - 74) = 4.04(x_2 - 7) + 4.36(x_1 - 6.8)$$

$$x_3 = 16.07 + 4.36x_1 + 4.04\,x_2$$

(b) The final grade of student who scored 9 and 7 marks is obtained by substituting $x_1 = 9$ and $x_2 = 7$ in the regression equation:

$$x_3 = 16.07 + 4.36(9) + 4.04(7) = 16.07 + 39.24 + 28.28 = 83.59 \text{ or } 84$$

Similarly, the final grade of students who scored 4 and 8 marks can also be obtained by substituting $x_1 = 4$ and $x_2 = 8$ in the regression equation:

$$x_3 = 16.07 + 4.36(4) + 4.04\,(8) = 16.07 + 17.44 + 32.32 = 65.83 \text{ or } 66$$

Example 15.3: The following data show the corresponding values of three variables x_1, x_2 and x_3. Find the least-square regression equation of x_3 on x_1 and x_2. Estimate x_3 when $x_1 = 10$ and $x_2 = 6$.

$\bar{x}_1$	:	3	5	6	8	12	14
$\bar{x}_2$	:	16	10	7	4	3	2
$\bar{x}_3$	:	90	72	54	42	30	12

Solution: The regression equation of x_3 on x_2 and x_1 can be written as follows:

$$x_3 - \bar{x}_3 = \left(\frac{r_{23} - r_{13}r_{12}}{1 - r_{12}^2}\right)\left(\frac{s_3}{s_2}\right)(x_2 - \bar{x}_2) + \left(\frac{r_{13} - r_{23}r_{12}}{1 - r_{12}^2}\right)\left(\frac{s_3}{s_1}\right)(x_1 - \bar{x}_1)$$

Calculations for regression equations are shown in the table below:

x_1	$(x_1 - \bar{x}_1) = X_1$	X_1^2	x_2	$(x_2 - \bar{x}_2) = X_2$	X_2^2	x_3	$(x_3 - \bar{x}_3) = X_3$	X_3^2	X_1X_2	X_1X_3	X_2X_3
3	−5	25	16	9	81	90	40	1600	−45	−200	360
5	−3	9	10	3	9	72	22	484	−9	−66	66
6	−2	4	7	0	0	54	4	16	0	−8	0
8	0	0	4	−3	9	42	−8	64	0	0	24
12	4	16	3	−4	16	30	−20	400	−16	−80	80
14	6	36	2	−5	25	12	−38	1444	−30	−228	190
48	0	90	42	0	140	300	0	4008	−100	−582	720

$$\bar{x}_1 = \frac{\Sigma x_1}{n} = \frac{48}{6} = 8;\ \bar{x}_2 = \frac{\Sigma x_2}{n} = \frac{42}{6} = 7;\ \bar{x}_3 = \frac{\Sigma x_3}{n} = \frac{300}{6} = 50$$

$$s_1 = \sqrt{\frac{\Sigma\,(x_1 - \bar{x}_1)^2}{n}} = \sqrt{\frac{90}{8}} = \sqrt{15} = 3.87$$

$$s_2 = \sqrt{\frac{\Sigma\,(x_2 - \bar{x}_2)^2}{n}} = \sqrt{\frac{140}{6}} = \sqrt{23.33} = 4.83$$

$$s_3 = \sqrt{\frac{\Sigma\,(x_3 - \bar{x}_3)^2}{n}} = \sqrt{\frac{4008}{6}} = \sqrt{668} = 25.85$$

$$r_{12} = \frac{\Sigma X_1X_2}{\sqrt{\Sigma X_1^2}\,\sqrt{\Sigma X_2^2}} = \frac{-100}{\sqrt{90 \times 140}} = -0.891$$

$$r_{13} = \frac{\Sigma X_1 X_3}{\sqrt{\Sigma X_1^2}\sqrt{\Sigma X_3^2}} = \frac{-582}{\sqrt{90 \times 4008}} = -0.969$$

$$r_{23} = \frac{\Sigma X_2 X_3}{\sqrt{\Sigma X_2^2}\sqrt{\Sigma X_3^2}} = \frac{720}{\sqrt{140 \times 4008}} = 0.961$$

Substituting values in the regression equation, we have

$$(x_3 - \bar{x}_3) = \left(\frac{r_{23} - r_{13}r_{12}}{1 - r_{12}^2}\right)\left(\frac{s_3}{s_2}\right)(x_2 - \bar{x}_2) + \left(\frac{r_{13} - r_{23}r_{12}}{1 - r_{12}^2}\right)\left(\frac{s_3}{s_1}\right)(x_1 - \bar{x}_1)$$

$$x_3 - 50 = \left[\frac{0.961 - (-0.969 \times -0.891)}{1 - (-0.891)^2}\right]\left(\frac{25.85}{4.83}\right)(x_2 - 7)$$

$$+ \left[\frac{-0.969 - (0.961 \times -0.891)}{1 - (-0.891)^2}\right]\left(\frac{25.85}{3.87}\right)(x_1 - 8)$$

$$x_3 - 50 = 2.546(x_2 - 7) - 3.664(x_1 - 8)$$

$$x_3 = 2.546x_2 - 3.664x_1 + 61.49$$

When $x_1 = 10$ and $x_2 = 6$, we have $x_3 = 2.546(6) - 3.664(10) + 61.49 = 15.276 - 36.64 + 61.49 = 40$ approx.

Self-practice Problems 15A

15.1 The estimated regression equation for a model involving two independent variables and 10 observations is as follows:

$$\hat{y} = 25.1724 + 0.3960x_1 + 0.5980x_2$$

(a) Interpret b_1 and b_2 in this estimated regression equation

(b) Estimate y when $x_1 = 150$ and $x_2 = 190$

15.2 Consider the linear regression model

$$y = a + b_1x_1 + b_2x_2$$

where y = demand for a product (in units)

x_1 = annual average price of the product (in ₹/unit)

$x_2 = 1/a$; a is the advertising expenditure (₹ in lakh)

The data for regression analysis is as follows:

Year	1	2	3	4	5	6	7
y	53.52	51.34	49.31	45.93	51.65	38.26	44.29
x_1	1.294	1.344	1.332	1.274	1.056	1.102	0.930
a	1.837	1.053	0.905	0.462	0.576	0.260	0.363

Write the least-squares prediction equation and interpret the b_1 and b_2 estimates.

15.3 In a three variable distribution, we have: $s_1 = 3, s_2 = 4, s_3 = 5, r_{23} = 0.4, r_{31} = 0.6, r_{12} = 0.7$

Determine the regression equation of x_1 on x_2 and x_3 if the variables are measured from their means.

15.4 The following constants are obtained from measurement on length in mm(x_1), volume in cc(x_2), and weight in gm (x_3) of 300 eggs:

$\bar{x}_1 = 55.95$, $s_1 = 2.26$, $r_{12} = 0.578$

$\bar{x}_2 = 51.48$, $s_2 = 4.39$, $r_{13} = 0.581$

$\bar{x}_3 = 56.03$, $s_3 = 4.41$, $r_{23} = 0.974$

Obtain the linear regression equation of egg weight on egg length and egg volume. Hence estimate the weight of an egg whose length is 58 mm and volume is 52.5 cc.

15.5 Given the following data:

x_1 :	20	25	15	20	26	24
x_2 :	3.2	6.5	2.0	0.5	4.5	1.5
x_3 :	4.0	5.2	7.5	2.5	3.4	1.5

(a) Obtain the least-squares equation to predict x_1 values from those of x_2 and x_3.

(b) Predict x_1 when $x_2 = 3.2$ and $x_3 = 3.0$.

15.6 The in-charge of an adult education centre in a town wants to know as to how happy and satisfied the adult education centre students are. The following four factors were studied to measure the degree of satisfaction:

x_1 = age at the time of completing education

x_2 = number of living children

x_3 = annual income

x_4 = average number of social activities per week

The multiple regression equation was determined to be

$$y = -20 + 0.04x_1 + 30x_2 + 0.04x_3 + 36.3x_4$$

(a) Calculate the satisfaction level for a person who passed out at the age of 45, has two living children, has an annual income of ₹12,000, and has only one social activity in a week.

(b) Interpret the value of $a = -20$.

(c) Would a person be more satisfied with an additional income of ₹2000?

15.7 Find multiple regression equations of x_1, x_2, x_3 from the data relating to three variables given below:

x_1 :	4	6	7	9	13	15
x_2 :	15	12	8	6	4	3
x_3 :	30	24	20	14	10	4

15.8 Data Given the following data:

$\bar{x}_1 = 6,\quad \bar{x}_2 = 7,\quad \bar{x}_3 = 8,$

$s_1 = 1,\quad s_2 = 1,\quad s_3 = 3,$

$r_{12} = 0.6,\quad r_{13} = 0.7,\quad r_{23} = 0.8,$

(a) Find the regression equation of x_3 on x_1 and x_2.

(b) Estimate the value of x_3 when $x_1 = 4$ and $x_2 = 5$

Hints and Answers

15.1 (a) $b_1 = 0.3960$; every unit increase in the value of x_1 accounted for an increase of 0.3960 in the value of y when the influence of x_2 is held constant.

$b_2 = 0.5980$; every unit increase in the value of x_2 accounted for an increase of 0.5980 in the value of y when influence of x_1 is held constant.

(b) $y = 25.1724 + 0.3960\,(150) + 0.5980\,(190)$
$= 198.1924$

15.2 $\hat{y} = 69.753 - 10.091x_1 - 5.334x_2$

15.3 $$b_{12.3} = \frac{r_{12} - r_{23}r_{13}}{1 - r_{23}^2}\left(\frac{s_1}{s_2}\right) = \frac{0.7 - 0.4 \times 0.6}{1 - (0.4)^2}\left(\frac{3}{4}\right) = 0.411$$

$$b_{13.2} = \frac{r_{13} - r_{12}r_{32}}{1 - r_{23}^2}\left(\frac{s_1}{s_3}\right) = \frac{0.6 - 0.7 \times 0.4}{1 - (0.4)^2}\left(\frac{3}{5}\right) = 0.229$$

Required regression equation is:

$$x_1 = b_{12.3}x_2 + b_{13.2}x_3 = 0.411x_2 + 0.229x_3$$

15.4 The regression equation of x_3 on x_1 and x_2 can be written as:

$$x_3 - \bar{x}_3 = \frac{r_{23} - r_{12}r_{13}}{1 - r_{12}^2}\left(\frac{s_3}{s_2}\right)(x_2 - \bar{x}_2) + \frac{r_{13} - r_{12}r_{23}}{1 - r_{12}^2}\left(\frac{s_3}{s_1}\right)(x_1 - \bar{x}_1)$$

$$x_3 - 56.03 = \left(\frac{0.974 - (0.581) \times 0.578)}{1 - (0.578)^2}\right)\left(\frac{4.41}{4.39}\right)(x_2 - 51.48) + \left(\frac{0.581 - (0.974 \times 0.578)}{1 - (0.581)^2}\right)\left(\frac{4.41}{2.26}\right)(x_2 - 55.95)$$

$$x_3 - 56.03 = \left(\frac{0.974 - 0.336}{1 - 0.334}\right)\left(\frac{4.41}{4.39}\right)(x_2 - 51.48) + \left(\frac{.581 - .563}{1 - 0.334}\right)\left(\frac{4.41}{2.26}\right)(x_2 - 55.95)$$

$$x_3 - 56.03 = 0.962(x_2 - 51.48) + 0.053(x_1 - 55.95)$$
$$x_3 = 3.54 + 0.053x_1 + 0.962x_2$$

When length, $x_1 = 58$ and volume, $x_2 = 52.5$, weight of the egg would be

$$x_3 = 3.54 + 0.053(58) + 0.962(52.5) = 3.54 + 3.074 + 50.50 = 57.119 \text{ gm.}$$

15.5 (a) The least-squares equation of x_1 on x_2 and x_3 is:

$$\hat{x}_1 = a + b_{12.3}x_2 + b_{13.2}x_3 = 21.925 - 0.481x_2 + 0.299x_3$$

(b) For $x_2 = 3.2$ and $x_3 = 3.0$, we have

$$\hat{x}_1 = 21.925 - 0.481\,(3.2) + 0.299\,(3.0) = 21.283$$

15.6 (a) $y = -20 + 0.04x_1 + 30x_2 + 0.04x_3 + 36.3x_4$
$= -20 + 0.04(45) + 30(2) + 0.04(12000) + 36.3\,(1)$
$= -20 + 1.8 + 60 + 480 + 36.3 = 558.10$

(b) $a = -20$ implies that in the absence of all variables the person will be negatively satisfied (dissatisfied).

(c) For $x_3 = 14{,}000$ (=12,000 + 2000), we have
$y = -20 + 0.04(45) + 30(2) + 0.04(14000) + 6.48.10 = 638.10$

15.7 Normal equations are

$6a + 48b_{12.3} + 102b_{13.2} = 54$

$48a + 449b_{12.3} + 1034b_{13.2} = 339$

$102a + 1034b_{12.3} + 2188b_{13.2} = 720$

The required regression equation is, $x_1 = 16.479 + 0.389x_2 - 0.623x_3$.

15.8 (a) The regression equation of x_3 on x_1 and x_2 is

$$x_3 - \bar{x}_3 = \frac{s_3}{s_2}\left[\frac{r_{23} - r_{12}r_{13}}{1 - r_{12}^2}\right](x_2 - \bar{x}_2) + \frac{s_3}{s_1}\left[\frac{r_{13} - r_{23}r_{12}}{1 - r_{12}^2}\right](x_1 - \bar{x}_1)$$

$$= \frac{3}{2}\left[\frac{8 - (0.6 \times 0.7)}{1 - (0.6)^2}\right](x_2 - 7) + \frac{3}{1}\left[\frac{0.7 - (0.8 \times 0.6)}{1 - (0.6)^2}\right](x_1 - 6)$$

$$x_3 - 8 = 1.5\left[\frac{0.8 - 0.42}{0.64}\right](x_2 - 7) + 3\left[\frac{0.7 - 0.48}{0.64}\right](x_1 - 6)$$

$$x_3 = -4.41 + 0.89x_2 + 1.03x_1$$

(b) The value of x_3 when $x_1 = 4$ and $x_2 = 5$ would be,
$x_3 = -4.41 + 0.89 \times 5 + 1.03 \times 4 = 4.16$

15.4 STANDARD ERROR OF ESTIMATE FOR A MULTIPLE REGRESSION

The following three types of variations need to be calculated to measure variation in the dependent variable, y in a multiple linear regression analysis:

- Total variation or total sum of squares deviation $\quad \text{SST} = \sum_{i=1}^{n} (y - \bar{y})^2$
- Explained variation resulting from regression relationship between x and y $\quad \text{SSR} = \sum_{i=1}^{n} (\hat{y} - \bar{y})^2$
- Unexplained variation resulting from sampling error $\quad \text{SSE} = \sum_{i=1}^{n} (y - \hat{y})^2$

The calculation of each of this sum of squares is associated with a certain number of degrees of freedom. SST has $n - 1$ degrees of freedom (n sample observations minus one due to fixed sample mean), SSR has k degrees of freedom (k independent variables associated with estimating the value of y), SSE has $n - (k + 1)$ degrees of freedom (n sample observations minus $k+1$ constants $a, b_1, b_2, ..., b_k$) because in a multiple regression we estimate k slope parameters $b_1, b_2, ..., b_k$ plus an intercept a from a data set containing n observations.

These three types of variations are related by the following equation:

$$\sum_{i=1}^{n} (y - \bar{y})^2 = \sum_{i=1}^{n} (\hat{y} - \bar{y})^2 + \sum_{i=1}^{n} (y - \hat{y})^2$$

$$\text{SST} = \text{SSR} + \text{SSE}$$

This implies that the total variation in y can be divided into two parts: explained variation and unexplained variation as shown in Fig. 15.2.

If the multiple regression equation fits the entire data perfectly (i.e., all observations in the data set fall on the regression line), then value of the dependent variable is assumed to estimated accurately and hence there is no error. But if it does not happen, then to know the degree of accuracy in the estimation of the value of dependent variable, y, a measure called *standard error of estimate* is used.

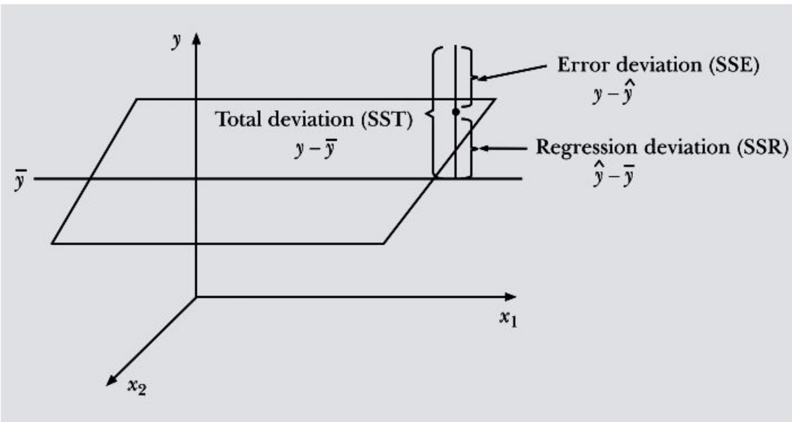


Figure 15.2 Decomposition of Total Variation

In multiple regression analysis, the sample standard error of the estimate represents the extent of variation (dispersion) of observations in the data set with respect to two independent variables or more independent variables.

15.4.1 Significance Test of Regression Model

To understand whether or not all independent variables, x_i, taken together significantly explain the variations in the dependent variable y. The F-test used to test the significance of the variations in the dependent variable y is based on the fact that at least one of the regression parameters in the regression equation must be zero. To apply F-test, the null hypothesis that all the regression parameters are zero is defined as

H_0: $b_1 = b_2 = \ldots = b_k = 0$ ← null hypothesis that y does not depend on all independent variables, x_i

H_1: at least one $b_i \neq 0$ ← alternative hypothesis that y depends on at least one of the independent variables, x_i

The analysis of multiple regression table for F-test statistic is shown in Table 15.2.

Table 15.2: ANOVA Table for a Multiple Regression

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	*F-Ratio*
Regression	SSR	k	$\text{MSR} = \dfrac{\text{SSR}}{k}$	$F = \dfrac{\text{MSR}}{\text{MSE}}$
Residual (or Error)	SSE	$n - (k+1)$	$\text{MSE} = \dfrac{\text{SSE}}{n-(k+1)}$	
Total	SST	$n - 1$		

Decision Rule

If the calculated value of Fcal is more than its table value at a given level of significance and degrees of freedom k for numerator and $n - k - 1$ for denominator, then H_0 is rejected. Hence, it may be concluded that the regression model has no significant prediction for the dependent variable. In other words, at least one of the independent variables is adding significant prediction for y.

If two variables are involved in the least-squares regression equation to predict the value of the dependent variable, then the *standard error of estimate* denoted by $S_{y.12}$ is given by

$$S_{y.12} = \sqrt{\frac{\text{SSE}}{n-3}} = \sqrt{\frac{\Sigma (y - \hat{y})^2}{n-3}} = \sqrt{\frac{\Sigma y^2 - b_1\Sigma x_1 y - b_2 \Sigma x_2 y}{n-3}} \tag{15-11}$$

The denominator in equation (15.11) indicates that in a multiple regression with two independent variables the standard error has $n - (2 + 1) = n - 3$ degrees of freedom (number of unrestricted chances for variation in the measurement). This is because the degrees of freedom is reduced from n to $2 + 1 = 3$ numerical constants a, b_1 and b_2 that have been estimated from the sample.

In general to determine $\hat{y}$ value involving $k + 1$ parameters $(a, b_1, b_2, \ldots, b_k)$ for the least-squares regression equation, $y = a + b_1x_1 + b_2x_2 + \ldots + b_kx_k$ with $n - (k + 1)$ residual degrees of freedom, the error term, e and standard error of estimate are computed as follows:

$$S^2_{y.12\ldots(k+1)} \text{ or } \sigma^2_e = \frac{\text{SSE}}{n-(k+1)} = \text{MSE} \quad \leftarrow \text{variance of error term } e \tag{15-12}$$

$$= \sqrt{\frac{\Sigma (y-\hat{y})^2}{n-(k+1)}} = \sqrt{\text{MSE}} \quad \leftarrow \text{standard error of estimate}$$

where y is sample values of dependent variable, $\hat{y}$ is the estimated values of dependent variable from the regression equation, n is the number of observations in the sample and k is the number of independent variables.

Thus, the standard error of estimate of y (say x_1) on x_2 and x_3 is defined as

$$S_{1.23} = \sqrt{\frac{(x_1 - \hat{x}_1)^2}{n-3}}$$

An alternative method of computing $S_{1.23}$ in terms of correlation coefficients r_{12}, r_{13}, and r_{23} is

$$S_{1.23} = s_1 \sqrt{\frac{1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}\, r_{13}\, r_{23}}{1 - r_{23}^2}}$$

By symmetry, we may write

$$S_{2.13} = s_2 \sqrt{\frac{1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}\, r_{13}\, r_{23}}{1 - r_{13}^2}}$$

$$S_{3.12} = s_3 \sqrt{\frac{1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}\, r_{13}\, r_{23}}{1 - r_{12}^2}}$$

15.5 COEFFICIENT OF MULTIPLE DETERMINATIONS

The coefficient of determination in multiple regressions denoted $R^2_{y.12}$ or $R^2_{1.23}$ is similar to the coefficient of determination r^2 in the linear regression. It represents the *proportion (fraction) of the total variation in the multiple values of dependent variable y, accounted for or explained by the independent variables in the multiple regression model.* The value of R^2 varies from zero to one.

The difference SST – SSE = SSR measures the variation in the sample values of the dependent variable y due to the changes (variations) among sample values of several independent variables in the least-squares equation. Thus, $R^2_{y.12}$ measures the ratio of the sum of squares due to regression to the total sum of squared deviations and is given by

$$R^2_{y.12} = \frac{SSR}{SST} = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{S^2_{y.12}}{S^2_y}$$

Adjusted (Corrected) R^2

The value of SST$(= S^2_y)$ remains same even if additional independent variables are added to the regression equation because it represents the sum of squares of the dependent variable. However, additional independent variables are likely to increase value of SSR, so value of R^2 is also likely to increase for additional independent variables.

An *adjusted* R^2 value takes into consideration: (i) the additional information which each additional independent variable brings to the regression analysis, and (ii) the changed degrees of freedom of SSE and SST.

In particular, the coefficient of multiple determinations as a measure of the proportion of total variation in the dependent variable x_1 accounted for or explained by the combined influence of the variations in the independent variables x_2 and x_3 can be defined as

$$R^2_{1.23} = 1 - \frac{S^2_{1.23}}{S^2_1}$$

where S^2_1 is the variance of the dependent variable x_1.

An adjusted coefficient of determination is defined as

$$\text{Adjusted } R^2_a = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)} = 1 - (1 - R^2)\frac{n-1}{n-(k+1)}$$

$$= 1 - \frac{MSE}{SST/(n-1)}$$

The decision to add an additional independent variable in a regression equation should suggest the increase in R^2 against the loss of one degree of freedom for error due to addition of the variable.

The three measures of performance of a regression model, (i) coefficient of determination R^2, (ii) mean square error, MSE, and (iii) adjusted coefficient of determination R^2_a can be obtained from ANOVA table as shown in Table 15.3.

Table 15.3: Measures of Performance of Regression Analysis

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	*F-ratio*
Regression	SSR	k	$MSR = \frac{SSR}{k}$	$F = \frac{MSR}{MSE}$
Residual error	SSE	$n-(k+1)$	$MSE = \frac{SSE}{n-(k+1)}$	
Total	SST	$n-1$		

15.6 MULTIPLE CORRELATION ANALYSIS

Multiple Regression Model: A mathematical equation that describes how the dependent variable, say y, is related to the independent variables $x_1, x_2, \ldots, x_k$ and a random error term e.

The **multiple correlation coefficient,** $R_{1.23} = \sqrt{R_{1.23}^2}$ measures the extent of association between a dependent variable x_1 and two independent variables x_2 and x_3 involved in the regression equation. In general, it is also being used to measures the extent of association between a dependent variable and several independent variables taken together.

The multiple correlation coefficients, $R_{y.12\ldots}$ is always measured as an absolute number because a few independent variables involved in the regression equation may have a negative (inverse) relationship with the dependent variable while the remaining may have a positive relationship.

The coefficient of multiple correlations can be expressed in terms of simple linear correlation coefficients r_{12}, r_{13}, and r_{23} as follows:

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}\,r_{13}\,r_{23}}{1 - r_{23}^2}}$$

By symmetry, we may also write

$$R_{2.13} = \sqrt{\frac{r_{21}^2 + r_{23}^2 - 2r_{12}\,r_{13}\,r_{23}}{1 - r_{13}^2}} \quad \text{and} \quad R_{3.12} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2r_{12}\,r_{13}\,r_{23}}{1 - r_{12}^2}}$$

The value of *multiple correlation coefficients* always lies between 0 and 1. The closer it is to 1, better is the linear relationship between the variables. If $R_{y.12} = 0$, then it implies no linear relationship between the variables and $R_{y.12} = 1$, implies that correlation is *perfect*.

15.7 PARTIAL CORRELATION ANALYSIS

Partial Coefficient of Correlation: It describes the relationship between one of the independent variables and the dependent variables, given that the other independent variables are held constant statistically.

The **partial correlation coefficient** measures the correlation between the dependent variable and one of the independent variables holding other independent variables constant rather than ignored in the analysis. If a dependent variable, x_1, and two independent variables, x_2 and x_3, are included in the partial correlation analysis, then the partial correlation between x_1 and x_2 holding x_3 constant is denoted by $r_{12.3}$. Similarly, the partial correlation between x_1 and x_3 holding x_2 constant is denoted by $r_{13.2}$. Depending upon the number of independent variables held constant, the partial correlation coefficient is called as zero-order, first-order and second-order correlation coefficients and so on.

The partial correlation coefficient between x_1 and x_2 keeping x_3 constant is determined as

$$r_{12.3}^2 = b_{12.3} \times b_{21.3} = \frac{r_{12} - r_{13}\,r_{23}}{1 - r_{23}^2}\left(\frac{s_1}{s_2}\right) \times \frac{r_{12} - r_{13}\,r_{23}}{1 - r_{13}^2}\left(\frac{s_2}{s_1}\right)$$

$$= \frac{(r_{12} - r_{13}\,r_{23})^2}{\left(1 - r_{23}^2\right)\left(1 - r_{13}^2\right)}$$

or $$r_{12.3} = \frac{r_{12} - r_{13}\,r_{23}}{\sqrt{1 - r_{13}^2}\,\sqrt{1 - r_{23}^2}}; \quad r_{13.2} = \frac{r_{13} - r_{12}\,r_{32}}{\sqrt{1 - r_{12}^2}\,\sqrt{1 - r_{32}^2}} \quad \text{and} \quad r_{23.1} = \frac{r_{23} - r_{21}\,r_{31}}{\sqrt{1 - r_{21}^2}\,\sqrt{1 - r_{31}^2}}$$

These results can also be obtained by applying the following formulae:

(a) $r_{12.3} = \sqrt{R^2_{12.3}} = \sqrt{1 - \frac{S^2_{1.23}}{S^2_{1.3}}}$, and $r_{13.2} = \sqrt{R^2_{13.2}} = \sqrt{1 - \frac{S^2_{1.23}}{S^2_{1.2}}}$

(b) $r_{12.3} = \sqrt{b_{12.3} \times b_{21.3}}$; $r_{13.2} = \sqrt{b_{13.2} \times b_{31.2}}$, and $r_{23.1} = \sqrt{b_{23.1} \times b_{32.1}}$

Limitations of Partial Correlation Analysis

- It is assumed that the simple (or zero order) correlation from which partial correlation is studied has linear relationship between the variables.
- The change in the value of dependent variable due to change in the value of independent variables is measured individually rather than jointly. However, in actual practice independent variables may be related to each other.
- The reliability of the partial correlation coefficient decreases as its order increases.

Standard Error of Partial Correlations

The reliability of the partial correlation coefficient in terms of standard error, $z = 1/\sqrt{n-3}$, depends on the number of independent variable held constant. For example, if one independent variable is held constant in the partial correlation, $r_{12.3}$, then standard error would be $z = 1/\sqrt{n-(3+1)} = 1/\sqrt{n-4}$ (one more degree of freedom is lost). Similarly, for partial correlation, $r_{12.34}$ in which two independent variables are held constant, then standard error would be $z = 1/\sqrt{n-(3+2)} = 1/\sqrt{n-5}$ (two more degrees of freedom are lost).

Example 15.4: Verify whether following partial correlation coefficients are significant.

(a) $r_{12.3} = 0.50;\ n = 29$

(b) $r_{12.34} = 0.60;\ n = 54$

Also set up the 95 per cent confidence limits of the correlation.

Solution: (a) Converting r into z, we get

$$z = \frac{1}{2}\log_e \frac{1+r}{1-r} = 1.1513 \log_{10}\frac{1+0.50}{1-0.50}$$
$$= 1.1513 \log 3 = 1.1513 \times 0.4771 = 0.549$$

The standard error, $z = \frac{1}{\sqrt{n-4}} = \frac{1}{\sqrt{29-4}} = \frac{1}{5} = 0.20$

The calculated value of z is more than 1.96 times of the standard error at $\alpha = 0.05$ level of significance. Hence, the correlation is significant. The 95 per cent confidence limits would be: $0.549 \pm (1.96 \times 0.2) = 0.157$ and 0.941

(b) Converting of r into z, we get

$$z = \frac{1}{2}\log_e \frac{1+r}{1-r} = 1.1513 \log_{10}\frac{1+0.60}{1-0.60} = 1.1513 \log 4$$
$$= 1.1513 \times 0.6021 = 0.693$$

The standard error of $z = \frac{1}{\sqrt{n-5}} = \frac{1}{\sqrt{54-5}} = \frac{1}{\sqrt{49}} = 0.143$

The calculated value of z is more than 1.96 times the standard error at $\alpha = 0.05$ level of significance. Hence, the correlation is significant. The 95 per cent confidence limits would be $0.693 \pm (1.96 \times 0.143) = 0.413$ and 0.973.

15.7.1 Relationship Between Multiple and Partial Correlation Coefficients

The results connecting multiple and partial correlation coefficients are as follows:

- $1 - R^2_{1.23} = (1 - r^2_{12})(1 - r^2_{13.2})$
  $1 - R^2_{2.13} = (1 - r^2_{21})(1 - r^2_{23.1})$
  $1 - R^2_{3.12} = (1 - r^2_{31})(1 - r^2_{32.1})$
- $1 - R_{1.234} = (1 - r^2_{12})(1 - r^2_{13.2})(1 - r^2_{14.23})$

Example 15.5: The correlation between a general intelligence test and school achievement in a group of children aged 6 to 15 years is 0.80. The correlation between the general intelligence test and age in the same group is 0.70 and the correlation between school achievement and age is 0.60. What is the correlation between general intelligence and school achievement in children of the same age? Comment on the result.

Solution: Let x_1 = general intelligence test; x_2 = school achievement; x_3 = age of children.

Given, $r_{12} = 0.8$, $r_{13} = 0.7$, and $r_{23} = 0.6$. Then the correlation between general intelligence test and school achievement, keeping the influence of age as constant, we have

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}} = \frac{0.8 - 0.7 \times 0.6}{\sqrt{1 - (0.7)^2}\sqrt{1 - (0.6)^2}}$$

$$= \frac{0.8 - 0.42}{\sqrt{0.51}\sqrt{0.64}} = \frac{0.38}{0.57} = 0.667$$

Hence, we conclude that intelligence and school achievement are associated to each other to the extent of $r_{12.3} = 0.667$ while the influence of the children's age is held constant.

Example 15.6: In a three variables distribution it is found that, $r_{12} = 0.70$, $r_{13} = 0.61$, and $r_{23} = 0.40$. Find the values of $r_{23.1}$, $r_{13.2}$, and $r_{12.3}$. *[Delhi Univ., M.Com, 2006]*

Solution: The partial correlation between variables 2 and 3 keeping the influence of variable 1 constant is given by

$$r_{23.1} = \frac{r_{23} - r_{12}\, r_{13}}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{13}^2}}$$

Substituting the given values, we get

$$r_{23.1} = \frac{0.40 - 0.70 \times 0.61}{\sqrt{1 - (0.70)^2}\sqrt{1 - (0.61)^2}} = \frac{0.40 - 0.427}{\sqrt{0.51}\sqrt{0.6279}}$$

$$= \frac{0.027}{0.714 \times 0.7924} = \frac{0.027}{0.5657} = 0.0477$$

Similarly, we get

$$r_{13.2} = \frac{r_{13} - r_{12} \bullet r_{23}}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{23}^2}} = \frac{0.61 - 0.70 \times 0.40}{\sqrt{1 - (0.70)^2}\sqrt{1 - (0.40)^2}} = 0.504$$

$$r_{12.3} = \frac{r_{12} - r_{13} \bullet r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}} = \frac{0.70 - 0.61 \times 0.40}{\sqrt{1 - (0.61)^2}\sqrt{1 - (0.40)^2}} = 0.633$$

Example 15.7: Based on the following data, calculate $R_{1.23}$, $R_{3.12}$ and $R_{2.13}$.

$\bar{x}_1 = 6.8$	$\bar{x}_2 = 7.0$	$\bar{x}_3 = 74$
$s_1 = 1.0$	$s_2 = 0.8$	$s_3 = 9$
$r_{12} = 0.6$	$r_{13} = 0.7$	$r_{23} = 0.65$

[HP Univ., MBA, 2005]

Solution: The coefficient of multiple determinations of two independent variables, 2 and 3, is given by

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}\, r_{13}\, r_{23}}{1 - r_{23}^2}}$$

Substituting the given values, we get

$$R_{1.23} = \sqrt{\frac{(0.6)^2 + (0.7)^2 - 2 \times 0.6 \times 0.7 \times 0.65}{1 - (0.65)^2}} = \sqrt{\frac{0.36 + 0.49 - 0.546}{0.5775}}$$

$$= \sqrt{0.526} = 0.725$$

$$R_{3.12} = \sqrt{\frac{r_{31}^2 + r_{23}^2 - 2r_{12}\, r_{13}\, r_{23}}{1 - r_{12}^2}} = \sqrt{\frac{(0.7)^2 + (0.65)^2 - 2 \times 0.6 \times 0.7 \times 0.65}{1 - (0.6)^2}}$$

$$= \sqrt{\frac{0.49 + 0.4225 - 0.546}{1 - 0.36}} = \sqrt{0.573} = 0.757$$

$$R_{2.13} = \sqrt{\frac{r_{12}^2 + r_{23}^2 - 2r_{12}\, r_{13}\, r_{23}}{1 - r_{13}^2}} = \sqrt{\frac{(0.6)^2 + (0.65)^2 - 2 \times 0.6 \times 0.7 \times 0.65}{1 - (0.7)^2}}$$

$$= \sqrt{\frac{0.36 + 0.4225 - 0.546}{0.51}} = \sqrt{0.464} = 0.681$$

Example 15.8: Suppose a computer has found, for a given set of values of variables x_1, x_2, and x_3 the correlation coefficients are: $r_{12} = 0.91$, $r_{13} = 0.33$, and $r_{23} = 0.81$. Explain whether these computations may be said to be free from errors.

[Madurai Kamaraj Univ., B.Com., 2007]

Solution: For determining whether the given computations are correct or not, we calculate the value of the partial correlation coefficient $r_{12.3}$ for variables 1 and 2 keeping the influence of variable 3 constant. If the value of $r_{12.3}$ is less than one, then the computations may be said to be free from errors.

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{1 - r_{13}^2}\,\sqrt{1 - r_{23}^2}} = \frac{0.91 - (0.33)(0.81)}{\sqrt{1 - (0.33)^2}\,\sqrt{1 - (0.81)^2}}$$

$$= \frac{0.91 - 0.2673}{\sqrt{1 - 0.1089}\,\sqrt{1 - 0.6561}} = \frac{0.6427}{\sqrt{0.8911 \times 0.3439}} = 1.161$$

Since the calculated values of $r_{12.3}$ are more than one, the computations given in the question are not free from errors.

Example 15.9: The simple correlation coefficients between temperature (x_1), yield of corn (x_2) and rainfall (x_3) are $r_{12} = 0.59$, $r_{13} = 0.46$ and $r_{23} = 0.77$. Calculate the partial correlation coefficient $r_{12.3}$ and multiple correlation coefficient $R_{1.23}$.

[Delhi Univ., M.Com; HP Univ., MBA, 2006]

Solution: Partial correlation coefficient $r_{12.3}$ is defined as

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{1 - r_{13}^2}\,\sqrt{1 - r_{23}^2}}$$

Substituting the given values, we get

$$r_{12.3} = \frac{0.59 - 0.46 \times 0.77}{\sqrt{1 - (0.46)^2}\,\sqrt{1 - (0.77)^2}} = \frac{0.59 - 0.3542}{\sqrt{1 - 0.2116}\,\sqrt{1 - 0.5529}}$$

$$= \frac{0.2358}{0.5665} = 0.416$$

Multiple correlation coefficients are defined as

$$R_{1.32} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}\, r_{13}\, r_{23}}{1 - r_{23}^2}} = \sqrt{\frac{(0.59)^2 + (0.46)^2 - 2\,(0.59 \times 0.46 \times 0.77)}{1 - (0.77)^2}}$$

$$= \sqrt{\frac{0.3481 + 0.2116 - 0.418}{0.4071}} = \sqrt{\frac{0.5597 - 0.418}{0.4071}}$$

$$= \sqrt{\frac{0.1417}{0.4071}} = 0.589$$

Example 15.10: A random sample of 5 years on the yield of a crop when observed for seed (x_1), rainfall (x_2) and temperature (x_3) revealed the following information:

$r_{12} = 0.80$	$r_{13} = -0.40$	$r_{23} = -0.56$
$s_1 = 4.42$	$s_2 = 1.10$	$s_3 = 8.50$

Calculate the following:

(a) Partial regression coefficient $b_{12.3}$ and $b_{13.2}$
(b) Standard error of estimate $S_{1.23}$
(c) Coefficient of multiple correlations $R_{1.23}$
(d) Coefficient of partial correlation $r_{12.3}$ between x_1 and x_2 holding x_3 constant

Solution: (a) Substituting the given values in the formulae for partial regression coefficients $b_{12.3}$ and $b_{13.2}$, we get

$$b_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{1 - r_{23}^2}\left(\frac{s_1}{s_2}\right) = \frac{0.80 - (-0.40)\,(-0.56)}{1 - (-0.56)^2}\left(\frac{4.42}{1.10}\right) = 3.370$$

$$b_{13.2} = \frac{r_{13} - r_{12}\, r_{23}}{1 - r_{23}^2}\left(\frac{s_1}{s_3}\right) = \frac{-0.40 - 0.80\,(-0.56)}{1 - (-0.56)^2}\left(\frac{4.42}{8.50}\right) = 0.036$$

(b) Standard error of estimate

$$S_{1.23} = s_1\sqrt{\frac{1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}\, r_{13}\, r_{23}}{1 - r_{23}^2}}$$

$$= 4.42\sqrt{\frac{1 - (0.8)^2 - (-0.4)^2 - (-0.56)^2 + 2\,(0.8) - (-0.4)\,(-0.56)}{1 - (-0.56)^2}}$$

$$= 4.42\sqrt{\frac{1 - 0.64 - 0.16 - 0.313 + 0.358}{0.6864}} = 2.642$$

(c) Coefficient of multiple correlations

$$r_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}\, r_{13}\, r_{23}}{1 - r_{23}^2}}$$

$$= \sqrt{\frac{(0.8)^2 + (-0.4)^2 - 2\,(0.8)\,(-0.4)\,(-0.56)}{1 - (-0.56)^2}} = 0.6433$$

(d) Coefficient of partial correlation, $r_{12.3}$

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{1 - r_{13}^2}\,\sqrt{1 - r_{23}^2}} = \frac{0.8 - (-0.40)\,(-0.56)}{\sqrt{1 - (-0.4)^2}\,\sqrt{1 - (-0.56)^2}}$$

$$= \frac{0.576}{0.759} = 0.758.$$

Example 15.11: On the basis of observations made on 35 cotton plants the total correlations of yield of cotton (x_1) and number of balls i.e. (the seed vessels) (x_2), and height (x_3) are found to be $r_{12} = 0.863$, $r_{13} = 0.648$ and $r_{23} = 0.709$. Determine the multiple correlation coefficient and partial correlation coefficient $r_{12.3}$ and $r_{13.2}$ and interpret your results.

Solution: A multiple correlation coefficient, $R_{1.23}$, is given by

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}\, r_{13}\, r_{23}}{1 - r_{23}^2}}$$

$$= \sqrt{\frac{(0.863)^2 + (0.648)^2 - 2\,(0.863)\,(0.648)\,(0.709)}{1 - (0.709)^2}}$$

$$= \sqrt{\frac{0.744 + 0.419 - 0.792}{0.498}} = \sqrt{\frac{0.371}{0.498}} = 0.863$$

The partial correlation coefficients, $r_{12.3}$ and $r_{13.2}$, are calculated as

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{1 - r_{13}^2}\,\sqrt{1 - r_{23}^2}} = \frac{0.863 - 0.648 \times 0.709}{\sqrt{1 - (0.648)^2}\,\sqrt{1 - (0.709)^2}} = \frac{0.409}{0.762 \times 0.705} = 0.761$$

$$r_{13.2} = \frac{r_{13} - r_{23}\, r_{12}}{\sqrt{1 - r_{23}^2}\,\sqrt{1 - r_{12}^2}} = \frac{0.648 - 0.709 \times 0.863}{\sqrt{1 - (0.709)^2}\,\sqrt{1 - (0.863)^2}} = \frac{0.037}{0.705 \times 0.505} = 0.103$$

Example 15.12: The following values have been obtained from the measurement of three variables x_1, x_2 and x_3.

$\bar{x}_1 = 6.80$	$\bar{x}_2 = 7.00$	$\bar{x}_3 = 7.40$
$s_1 = 1.00$	$s_2 = 0.80$	$s_3 = 0.90$
$r_{12} = 0.60$	$r_{13} = 0.70$	$r_{23} = 0.65$

(a) Obtain the regression equation of x_1 on x_2 and x_3

(b) Estimate the value of x_1 for $x_2 = 10$ and $x_2 = 9$

(c) Find the coefficient of a multiple determination $R_{1.23}^2$ from r_{12} and $r_{13.2}$

Solution: (a) The regression equation of x_1 on x_2 and x_3 is given by

$$(x_1 - \bar{x}_1) = \left(\frac{r_{12} - r_{13}\, r_{23}}{1 - r_{23}^2}\right)\left(\frac{s_1}{s_2}\right)(x_2 - \bar{x}_2) + \left(\frac{r_{13} - r_{12}\, r_{23}}{1 - r_{23}^2}\right)\left(\frac{s_1}{s_3}\right)(x_3 - \bar{x}_3)$$

$$x_1 - 6.8 = \left(\frac{0.60 - 0.70 \times 0.65}{1 - (0.65)^2}\right)\left(\frac{1}{0.80}\right)(x_2 - 7) + \left(\frac{0.70 - 0.60 \times 0.65}{1 - 10.657^2}\left(\frac{1}{0.90}\right)(x_3 - 7.4)\right)$$

$$x_1 - 6.8 = \left(\frac{0.60 - 0.455}{0.578}\right)(1.25)\,(x_2 - 7) + \left(\frac{0.70 - 0.39}{0.578}\right)(1.111)\,(x_3 - 7.4)$$

$$= 0.181(x_2 - 7) + 0.595(x_3 - 7.4)$$

or $\quad x_1 = 5.670 + 0.181x_2 + 0.595x_3$

(b) Substituting $x_2 = 10$ and $x_2 = 9$ in the regression equation obtained in part (a), we have

$$x_1 = 5.670 + 0.181(10) + 0.595(9) = 12.835$$

(c) Multiple and partial correlation coefficients are related as

$$R_{1.23}^2 = 1 - (1 - r_{12}^2)\,(1 - r_{13.2}^2)$$

$$r_{13.2} = \frac{r_{13} - r_{12}\, r_{23}}{\sqrt{1 - r_{12}^2}\,\sqrt{1 - r_{23}^2}} = \frac{0.70 - 0.60 \times 0.65}{\sqrt{1 - (0.60)^2}\,\sqrt{1 - (0.65)^2}}$$

$$= \frac{0.70 - 0.39}{0.8 \times 0.760} = 0.509$$

or $\quad r_{13.2}^2 = 0.259$

Substituting values for r_{12}^2 and $r_{13.2}^2$ for $R_{1.23}^2$, we have

$$R_{1.23}^2 = 1 - (1 - 0.36)\,(1 - 0.259) = 0.526.$$

Example 15.13: Given the following data, determine the regression equation of

(a) x_1 on x_2 and x_3

(b) x_2 on x_1 and x_3

$$r_{12} = 0.8; \quad r_{13} = 0.6; \quad r_{23} = 0.5;$$
$$\sigma_1 = 10; \quad \sigma_2 = 8; \quad \sigma_3 = 10.$$

[Punjab Univ., M.Com., 2006]

Solution: (a) Regression equations of x_1 on x_2 and x_3 is given by

$$x_1 = b_{12.3}\, x_2 + b_{13.2}\, x_3$$
$$= 0.833\, x_2 + 0.533\, x_3.$$

where $$b_{12.3} = \frac{\sigma_1}{\sigma_2} \times \frac{r_{12} - r_{13}\, r_{23}}{1 - r_{23}^2} = \frac{10}{8} \times \frac{0.8 - (0.6)\,(0.5)}{1 - (0.5)^2} = 0.833\,;$$

$$b_{13.2} = \frac{\sigma_1}{\sigma_3} \times \frac{r_{13} - r_{12}\, r_{23}}{1 - r_{23}^2} = \frac{10}{5} \times \frac{0.6 - (0.8)\,(0.5)}{1 - (0.5)^2} = 0.533.$$

(b) Regression equation of x_2 on x_1 and x_3

$$x_2 = b_{21.3}\, x_1 + b_{23.1}\, x_3$$
$$= 0.625\, x_1 + 0.05\, x_3.$$

where $$b_{21.3} = \frac{\sigma_2}{\sigma_1} \times \frac{r_{12} - r_{23}\, r_{13}}{1 - r_{13}^2} = \frac{8}{10} \times \frac{0.8 - (0.5)\,(0.6)}{1 - (0.6)^2} = 0.625;$$

$$b_{23.1} = \frac{\sigma_2}{\sigma_3} \times \frac{r_{23} - r_{12}\, r_{13}}{1 - r_{13}^2} = \frac{8}{5} \times \frac{0.5 - (0.8)\,(0.6)}{1 - (0.6)^4} = 0.05.$$

Example 15.14: In a trivarite distribution,

$$\sigma_1 = 3; \quad \sigma_2 = \sigma_3 = 5;$$
$$r_{12} = 0.6; \quad r_{23} = r_{31} = 0.8.$$

Find (a) $r_{23.1}$ and (b) $R_{1.23}$.

Solution: (a) Since $$r_{23.1} = \frac{r_{23} - r_{12}\, r_{13}}{\sqrt{1 - r_{12}^2}\,\sqrt{1 - r_{13}^2}} = \frac{0.8 - 0.6 \times 0.8}{\sqrt{1 - (0.6)^2}\,\sqrt{1 - (0.8)^2}} = \frac{0.8 - 0.48}{\sqrt{0.64}\,\sqrt{0.36}} = 0.667$$

(b) $$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2\, r_{12}\, r_{13}\, r_{23}}{1 - r_{23}^2}} = \sqrt{\frac{(0.6)^2 + (0.8)^2 - 2\,(0.6)\,(0.8)\,(0.8)}{1 - (0.8)^2}}$$

$$= \sqrt{\frac{0.36 + 0.64 - 0.768}{0.36}} = \sqrt{0.644} = 0.803.$$

Example 15.15: A teacher wishes to determine the relationship of grades on a final examination to grades on two quizes given during the semester. Based on x_1, x_2, and x_3 the grades of a student on the first quiz, second quiz, and final examination, respectively, he made the following computations for a total of 120 students:

$$\bar{x}_1 = 6.8; \quad \bar{x}_2 = 7; \quad \bar{x}_3 = 74$$
$$s_1 = 1.0; \quad s_2 = 0.80; \quad s_3 = 9.0$$
$$r_{12} = 0.60; \quad r_{13} = 0.70; \quad r_{23} = 0.65.$$

(a) Find the least square regression equation of x_3 on x_1 and x_2.

(b) Estimate the final grades to two students who scored, resepectively, 9 and 7, 4 and 5 in the two quizes. *[HP Univ., MBA, 2000]*

Solution: (a) The regression equation of x_3 on x_1 and x_2.

$$x_3 - \bar{x}_3 = \left(\frac{r_{23} - r_{12} r_{13}}{1 - r_{12}^2}\right)\left(\frac{s_3}{s_2}\right)(x_2 - \bar{x}_2) + \left(\frac{r_{13} - r_{23} r_{12}}{1 - r_{12}^2}\right)\left(\frac{s_3}{s_1}\right)(x_1 - \bar{x}_1)$$

$$x_3 - 74 = \left(\frac{0.65 - 0.7 \times 0.6}{1 - (0.6)^2}\right)\left(\frac{9}{0.8}\right)(x_2 - 7) + \left(\frac{0.7 - 0.65 \times 0.6}{1 - (0.6)^2}\right)\left(\frac{9}{1}\right)(x_1 - 6.8)$$

$$= \left(\frac{0.65 - 0.42}{0.64}\right)\left(\frac{9}{0.8}\right)(x_2 - 7) + \left(\frac{0.7 - 0.39}{0.64}\right)(9)(x_1 - 6.8)$$

$$= 4.04\,(x_2 - 7) + 4.36\,(x_1 - 6.8)$$

$$= 4.04\,x_2 - 28.28 + 4.36\,x_1 - 29.65$$

$$x_3 = 16.07 + 4.36\,x_1 + 4.04\,x_2.$$

(b) Final grades of students who scored $x_1 = 9$ and $x_2 = 7$ marks are

$$x_3 = 16.07 + 4.36\,(9) + 4.04\,(7) = 83.59 \text{ or } 84.$$

Final grades of students who scored $x_1 = 4$ and $x_2 = 5$ marks are

$$x_2 = 16.04 + 4.36\,(4) + 4.04\,(5) = 53.68 \text{ or } 54.$$

Example 15.16: Calculate (a) $R_{1.23}$, (b) $R_{3.12}$, and (c) $R_{2.13}$ for the following data:

$\bar{x}_1 = 6.8$; $\bar{x}_2 = 7.0$; $\bar{x}_3 = 74$

$s_1 = 1.0$; $s_2 = 0.8$; $s_3 = 9$

$r_{12} = 0.6$; $r_{13} = 0.7$; $r_{23} = 0.65$. [*HP, Univ., MBA, 2005*]

Solution: (a) $R_{1.23} = \sqrt{\dfrac{r_{12}^2 + r_{13}^2 - 2\,r_{12}\,r_{13}\,r_{23}}{1 - r_{23}^2}} = \sqrt{\dfrac{(0.6)^2 + (0.7)^2 - 2\times 0.6\times 0.7\times 0.65}{1-(0.65)^2}}$

$$= \sqrt{\frac{0.36 + 0.49 - 0.546}{0.5776}} = \sqrt{0.526} = 0.725.$$

$$R_{3.12} = \sqrt{\frac{r_{13}^2 + r_{23}^2 - 2\,r_{12}\,r_{13}\,r_{23}}{1 - r_{12}^2}} = \sqrt{\frac{(0.7)^2 + (0.65)^2 - 2\times 0.6\times 0.7\times 0.65}{1-(0.6)^2}}$$

$$= \sqrt{\frac{0.49 + 0.4225 - 0.546}{1 - 0.36}} = \sqrt{0.573} = 0.757.$$

$$R_{2.13} = \sqrt{\frac{r_{12}^2 + r_{23}^2 - 2\,r_{12}\,r_{13}\,r_{23}}{1 - r_{13}^2}} = \sqrt{\frac{(0.6)^2 + (0.65)^2 - 2\times 0.6\times 0.7\times 0.65}{1-(0.7)^2}}$$

$$= \sqrt{\frac{0.36 + 0.4225 - 0.546}{0.51}} = \sqrt{0.464} = 0.681.$$

Example 15.17: The following constants are obtained from measurements on length (x_1) in mm, volume (x_2) in cc, and weight (x_3) in g of 300 eggs:

$\bar{x}_1 = 55.95$; $s_1 = 2.26$; $r_{12} = 0.578$

$\bar{x}_2 = 51.48$; $s_2 = 4.39$; $r_{13} = 0.581$

$\bar{x}_3 = 56.03$; $s_3 = 4.41$; $r_{23} = 0.974$.

Obtain the linear regression equation of egg weight on egg length and volume. Hence, estimate the weight of an egg whose length is 58 mm and volume is 52.5 cc.

Solution: The linear regression equation of egg weight on egg length and volume, that is, x_3 on x_1 and x_2 is given by

$$x_3 - \bar{x}_3 = \left(\frac{r_{23} - r_{13}r_{12}}{1 - r_{12}^2}\right)\left(\frac{s_3}{s_2}\right)(x_2 - \bar{x}_2) + \left(\frac{r_{13} - r_{23}\,r_{12}}{1 - r_{12}^2}\right)\left(\frac{s_3}{s_1}\right)(x_1 - \bar{x}_1)$$

$$x_3 - 56.03 = \left[\frac{0.974 - 0.581\times 0.578}{1-(0.578)^2}\right]\left(\frac{4.41}{4.39}\right)(x_2 - 51.48)$$

$$+ \left[\frac{0.581 - 0.974\times 0.578}{1-(0.578)^2}\right]\left(\frac{4.41}{2.26}\right)(x_1 - 55.95)$$

$$= \left(\frac{0.974 - 0.335}{1 - 0.334}\right)\left(\frac{4.41}{4.39}\right)(x_2 - 51.48) + \left(\frac{0.581 - 0.563}{1 - 0.334}\right)\left(\frac{4.41}{2.26}\right)(x_1 - 55.95)$$

$$= 0.964\,(x_2 - 51.48)\,0.053\,(x_1 - 55.95)$$
$$= 0.964\,x_2 - 49.63 + 0.053\,x_1 - 2.97$$
$$x_3 = 3.43 + 0.053\,x_1 + 0.964\,x_2.$$

If length, $x_1 = 58$ and volume, $x_2 = 52.5$, the weight of the egg would be

$$x_3 = 3.43 + 0.053\,(58) + 0.964\,(52.5) = 57.11 \text{ g}.$$

Example 15.18: In a trivariate distribution: $\sigma_1 = 3$, $\sigma_2 = 4$, $\sigma_3 = 5$; $r_{23} = 0.4$, $r_{13} = 0.6$, $r_{12} = 0.7$. If variables are measured from their mean, then determine the regression equation of x_1 on x_2 and x_3. *[Annamalai Univ., M.Com., 2003]*

Solution: The required regression equation of x_1 on x_2 and x_3 is given by

$$x_1 = b_{12.3}x_2 + b_{13.2}x_3$$
$$= 0.410\,x_2 + 0.229\,x_3.$$

where $b_{12.3} = \dfrac{\sigma_1}{\sigma_2} \times \dfrac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2} = \dfrac{3}{4} \times \dfrac{0.7 - 0.6 \times 0.4}{1 - (0.4)^2} = 0.410;$

$$b_{13.2} = \frac{\sigma_1}{\sigma_3} \times \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2} = \frac{3}{5} \times \frac{0.6 - 0.7 \times 0.4}{1 - (0.4)^2} = 0.229.$$

Example 15.19: (a) The simple correlation coefficient between temperature (x_1), corn yield (x_2), and rainfall (x_3) are $r_{12} = 0.59$, $r_{13} = 0.46$, and $r_{23} = 0.77$. Calculate partial correlation coefficient $r_{12.3}$ and multiple correlation coefficient $R_{1.23}$. *[HP Univ., M.Com., 2005]*

Solution: (a) $r_{12.3} = \dfrac{r_{12} - r_{13}\,r_{23}}{\sqrt{1 - r_{13}^2}\,\sqrt{1 - r_{23}^2}} = \dfrac{0.59 - 0.46 \times 0.77}{\sqrt{1 - (0.46)^2}\,\sqrt{1 - (0.77)^2}}$

$$= \frac{0.59 - 0.3542}{\sqrt{1 - 0.2116}\,\sqrt{1 - 0.5529}} = \frac{2358}{0.5940} = 0.397.$$

(b) $R_{1.23} = \dfrac{\sqrt{r_{12}^2 + r_{13}^2 - 2\,r_{12}\,r_{13}\,r_{23}}}{1 - r_{23}^2} = \dfrac{\sqrt{(0.59)^2 + (0.46)^2 - 2(0.59 \times 0.46 \times 0.77)}}{1 - (0.77)^2}$

$$= \frac{\sqrt{0.3481 + 0.2116 - 0.418}}{0.4071} = \frac{\sqrt{0.1417}}{0.4071} = 0.589.$$

Conceptual Questions 15A

1. What is the relationship between a residual and the standard error of estimate? If all residuals about a regression line are zero, what is the value of the standard error of estimate?
2. What information is provided by the variance computed about the regression line by the standard error of estimate?
3. What are 'normal equations' and how are they used in a multiple regression analysis?
4. Define the following terms
   (a) Standard error of estimate,
   (b) Coefficient of multiple determination, and
   (c) Coefficient of partial and multiple correlations.
5. Explain the objectives of performing a multiple regression and a correlation analysis?
6. Why may it be of interest to perform tests of hypotheses in multiple regression problems?
7. Define partial and multiple correlations. With the help of an example distinguish between partial and multiple correlations.
8. What are multiple linear regressions? Explain the difference between simple linear and multiple linear regressions.
9. Explain the concept of multiple regressions and try to find out an example in the practical field where multiple regression analysis is likely to be helpful.
10. Distinguish between partial and multiple correlations and point out their usefulness in statistical analysis.
11. (a) In the multiple regression equation of x_1 on x_2 and x_3, what are the two progression coefficients and how do you interpret them?

(b) Explain the concepts of simple, partial and multiple correlations.

(c) When is a multiple regression needed? Explain with the help of an example.

12. Under what conditions is it important to use the adjusted multiple coefficient of determination?

13. Can you judge how well a regression model fits the data by considering the mean square error only. Explain.

14. Explain why the multiple coefficient of determination never decreases as variables are added to the multiple regression equation.

Self-practice Problems 15B

15.9 In a three variables distribution, it is found that $r_{12} = 0.7$, $r_{13} = 0.61$, and $r_{23} = 0.4$. Find the values of the partial correlation coefficients $r_{12.3}$, $r_{23.1}$, and $r_{13.2}$.

15.10 The following data present the values of the dependent variable y and the two independent variables x_1 and x_2:

y :	6	8	9	11	12	14
x_1 :	14	16	17	18	20	23
x_2 :	21	22	27	29	31	32

Compute the following:

(a) Multiple regression coefficients $b_{12.3}$, $b_{13.2}$, $b_{23.1}$

(b) A multiple correlation coefficient $R_{1.23}$

(c) Partial correlation coefficients $r_{12.3}$, $r_{13.2}$, and $r_{23.1}$

(d) Standard error of estimate $S_{1.23}$

15.11 For a given set of values of x_1, x_2, and x_3, the computer has found that $r_{12} = 0.96$, $r_{13} = 0.36$, and $r_{23} = 0.78$. Examine whether these computations may be said to be free from errors.

15.12 The simple correlation coefficients between variable x_1, x_2, and x_3 are, respectively, $r_{12} = 0.41$, $r_{13} = 0.71$, and $r_{23} = 0.50$. Calculate the partial correlation coefficients $r_{12.3}$, $r_{13.2}$, and $r_{23.1}$.

15.13 In a trivariate distribution, it is found that $r_{12} = 0.8$, $r_{13} = 0.4$, and $r_{23} = 0.56$. Find the value of $r_{23.2}$, $r_{13.2}$, and $r_{23.1}$. *[Madras Univ., M Com, 2008]*

15.14 The simple correlation coefficients between profits (x_1), sales (x_2), and advertising expenditure (x_3) of a factory are $r_{12} = 0.69$, $r_{13} = 0.45$, and $r_{23} = 0.58$. Find the partial correlation coefficients $r_{12.3}$, and $r_{13.3}$ and interpret them.

15.15 (a) Simple coefficients of correlation between two variables out of three are as follows:

$$r_{12} = 0.8;\ r_{13} = 0.7, \text{ and } r_{23} = 0.6$$

Find the partial coefficients of correlation, $r_{12.3}$, $r_{13.2}$ and $r_{23.1}$

(b) If $r_{12} = 0.86$, $r_{13} = 0.65$ and $r_{23} = 0.72$, then prove that $r_{12.3} = 0.743$

15.16 On the basis of observations made on 30 cotton plants, the total correlation of yield of cotton (x_1), the number of balls, i.e. seed vessels (x_2) and height (x_3) are found to be;

$$r_{12} = 0.80;\ r_{13} = 0.65, \text{ and } r_{23} = 0.70$$

Compute the partial correlation between yield of cotton and the number of balls, eliminating the effect of height.

15.17 The following simple correlation coefficients are given:

$$r_{12} = 0.98,\ r_{13} = 0.44, \text{ and } r_{23} = 0.54$$

Calculate the partial coefficient of correlation between first and third variables keeping the effect of second variable constant.

15.18 (a) Do you find the following data consistent:

$$r_{12} = 0.07,\ r_{13} = -0.6, \text{ and } r_{23} = 0.90$$

(b) The simple correlation coefficient between temperature (x_1), yield (x_2) and rainfall (x_3) are $r_{12} = 0.6$, $r_{13} = 0.5$ and $r_{23} = 0.8$. Determine a multiple correlation $R_{1.23}$.

15.19 (a) The following zero order correlation coefficient are given as

$$r_{12} = 0.98,\ r_{13} = 0.44, \text{ and } r_{23} = 0.54$$

Calculate a multiple correlation coefficient treating first variable as dependent and second and third variables as independents.

(b) The following zero order correlation coefficients are given: $r_{12} = 0.5$, $r_{13} = 0.6$ and $r_{23} = .7$

Calculate the multiple correlation coefficients $R_{1.23}$, $R_{2.13}$ and $R_{3.12}$

Hints and Answers

15.9
$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{1-r_{13}^2}\,\sqrt{1-r_{23}^2}} = \frac{0.70 - 0.60 \times 0.40}{\sqrt{1-(0.60)^2}\,\sqrt{1-(0.40)^2}}$$
$$= 0.633$$

Similarly $r_{23.1} = 0.049$ and $r_{13.2} = 0.504$

15.10 (a) $b_{12.3} = 0.057$, $b_{13.2} = 0.230$; $b_{23.1} = 0.093$

(b) $R_{1.23} = 0.99$

(c) $r_{12.3} = 0.88$; $r_{13.2} = 0.78$, $r_{23.1} = 0.28$

(d) $S_{1.23} = 0.34$

15.11 Since $r_{12.3} = 1.163$ (> 1), the given computations are not free from error.

15.12 $r_{12.3} = 0.09$, $r_{23.1} = 0.325$ and $r_{13.2} = 0.639$

15.13 $r_{12.3} = 0.759$; $r_{23.1} = 0.436$ and $r_{13.2} = -0.097$

15.14 $r_{12.3} = 0.589$; $r_{13.2} = 0.085$

15.15 (a)
$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}} = \frac{0.8-(0.7\times 0.6)}{\sqrt{1-(0.7)^2}\sqrt{1-(0.6)^2}}$$
$$= \frac{0.8-0.42}{0.714\times 0.8} = \frac{0.38}{0.5712} = 0.665$$
$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{23}^2}} = \frac{0.7-(0.8\times 0.6)}{\sqrt{1-(0.8)^2}\sqrt{1-(0.6)^2}}$$
$$= \frac{0.7-0.48}{0.6\times 0.8} = \frac{0.22}{0.48} = 0.458$$
$$r_{23.1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{13}^2}} = \frac{0.6-(0.8\times 0.7)}{\sqrt{1-(0.8)^2}\sqrt{1-(0.7)^2}}$$
$$= \frac{0.6-0.56}{0.6\times 0.49} = \frac{.04}{0.4284} = 0.093$$

(b)
$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}}} = \frac{0.86-(0.65\times 0.72)}{\sqrt{1-(0.65)^2}\sqrt{1-(0.72)^2}}$$
$$= \frac{0.86-0.468}{\sqrt{1-0.4225}\sqrt{1-0.5184}}$$
$$= \frac{0.392}{0.7599\times 0.6939} = \frac{0.392}{0.527} = 0.743$$

15.16
$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}} = \frac{0.8-(0.65\times 0.7)}{\sqrt{1-(0.65)^2}\sqrt{1-(0.7)^2}}$$
$$= \frac{0.8-0.455}{\sqrt{1-0.4225}\sqrt{1-0.49}} = \frac{0.345}{0.7599\times 0.7141}$$
$$= \frac{0.345}{0.5426} = 0.6357$$

15.17
$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{23}^2}} = \frac{0.44-(0.98\times 0.54)}{\sqrt{1-(0.98)^2}\sqrt{1-(0.54)^2}}$$
$$= \frac{0.44-0.5292}{\sqrt{1-0.9604}\sqrt{1-0.2916}} = \frac{-0.0892}{0.199\times 0.842}$$
$$= \frac{-0.0892}{0.1676} = -0.5322$$

15.18 (a)
$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}} = \frac{7-(-0.6\times 0.9)}{\sqrt{1-(-0.6)^2}\sqrt{1-(0.9)^2}}$$
$$= \frac{0.7+0.54}{\sqrt{1-0.36}\sqrt{1-0.81}} = \frac{1.24}{0.8\times 0.436} = 3.56$$

Since the value of $r_{12.3}$ is 3.56, the data are inconsistent as the value of any partial coefficient of correlation cannot exceed unity or 1.

(b)
$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1-r_{23}^2}}$$
$$= \sqrt{\frac{(0.6)^2 + (0.5)^2 - 2\times 0.6\times 0.5\times 0.8}{1-(0.8)^2}}$$
$$= \sqrt{\frac{0.36+0.25-0.48}{1-0.64}} = 0.6$$

15.19 (a)
$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1-r_{23}^2}}$$
$$= \sqrt{\frac{(0.98)^2 + (0.44)^2 - 2\times 0.98\times 0.44\times 0.54}{1-(0.54)^2}}$$
$$= \sqrt{\frac{0.9604+0.1936-0.4657}{1-0.2916}} = 0.986$$

(b)
$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1-r_{23}^2}}$$
$$= \sqrt{\frac{(0.5)^2 + (0.6)^2 - 2\times 0.5\times 0.6\times 0.7}{1-(0.7)^2}} = 0.622$$
$$R_{2.13} = \sqrt{\frac{r_{21}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}}{1-r_{13}^2}}$$
$$= \sqrt{\frac{(0.5)^2 + (0.7)^2 - 2\times 0.5\times 0.6\times 0.7}{1-(0.6)^2}} = 0.707$$
$$R_{3.13} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2r_{12}r_{13}r_{23}}{1-r_{12}^2}}$$
$$= \sqrt{\frac{(0.6)^2 + (0.7)^2 - 2\times 0.5\times 0.6\times 0.7}{1-(0.5)^2}} = 0.757$$

14. If a multiple correlation coefficient $R_{1.23} = 1$, then $R_{2.13}$ is
(a) 0 (b) −1
(c) 1 (d) none of these

15. If a multiple correlation coefficient $R_{1.23} = 1$, then it implies a
(a) lack of linear relationship
(b) perfect relationship
(c) reasonably good relationship
(d) none of these

16. Which of following relationship is true?
(a) $R_{1.23} \leq r_{12}$ (b) $R_{1.23} \geq r_{12}$
(c) $R_{1.23} = r_{12}$ (d) $R_{1.23} \geq -1$

17. In the regression equation $y = a + b_1x_1 + b_2x_2$, y is independent of x when
(a) $b_2 = 0$ (b) $b_2 = 1$
(c) $b_2 = -1$ (d) none of these

18. Since $r^2 = 1 - \{(y - \hat{y})^2/(y - \bar{y})^2\}$, then r^2 is equal to
(a) 1 – SSR/SST (b) 1 – SSE/SSR
(c) 1 – SST/SSE (d) 1 – SSE/SST

19. In regression analysis, the explained deviation of the dependent variable y is given by
(a) $\Sigma (y - \bar{y})^2$ (b) $\Sigma (\hat{y} - \bar{y})$
(c) $\Sigma (y - \bar{y})^2$ (d) none of these

20. The relationship between the multiple correlation coefficient of x_1 on x_2 and x_3 and the standard error of estimate is given by the expression
(a) $R_{1.23} = \sqrt{1 - \frac{S_{1.23}^2}{S_1^2}}$ (b) $R_{1.23} = \sqrt{1 - \frac{S_{1.23}^2}{S_2^2}}$
(c) $R_{1.23} = \sqrt{1 - \frac{S_{1.23}^2}{S_3^2}}$ (d) none of these

21. Which of the following relationship is true
(a) $r_{12.3} = \sqrt{b_{12.3} \times b_{21.3}}$ (b) $r_{12.3} = \sqrt{b_{13.2} \times b_{31.2}}$
(c) $r_{12.3} = \sqrt{b_{23.1} \times b_{32.1}}$ (c) all of these

22. The coefficient of determination in a multiple regression is given by
(a) $R^2_{y.12} = 1 - (SSR/SST)$
(b) $R^2_{y.12} = 1 - (SSE/SSR)$
(c) $R^2_{y.12} = 1 - (SST/SSE)$
(d) $R^2_{y.12} = 1 - (SSE/SST)$

23. The standard error of estimate involved to predict the value dependent variable is given by
(a) $S_{y.12} = \sqrt{SSE/n-2}$ (b) $S_{y.12} = \sqrt{SSE/n-3}$
(c) $S_{y.12} = \sqrt{SSR/n-2}$ (d) $S_{y.12} = \sqrt{SSR/n-3}$

24. Adjusted a multiple coefficient of determination R_a^2 is given by
(a) $R_a^2 = 1 - \frac{MSE}{SST/(n-1)}$ (b) $R_a^2 = \frac{MSE}{SSE/(n-1)}$
(c) $R_a^2 = 1 - \frac{MSR}{SSE/(n-1)}$ (d) $R_a^2 = \frac{MSR}{SSE/(n-1)}$

25. The F-ratio used in testing for the existing of a regression relationship between a dependent variable and any of the independent variable is given by
(a) $F = \frac{R^2}{1-R^2}\left[\frac{n-(k+1)}{n}\right]$
(b) $F = \frac{R^2}{1-R^2}\left[\frac{n+(k-1)}{n}\right]$
(c) $F = \frac{R^2}{1-R^2}\left[\frac{n-(k+1)}{k}\right]$
(d) $F = \frac{R^2}{1-R^2}\left[\frac{n+(k-1)}{k}\right]$

Concepts Quiz Answers

1. T	**2.** T	**3.** F	**4.** T	**5.** T	**6.** F	**7.** T	**8.** T	**9.** T
10. F	**11.** (c)	**12.** (b)	**13.** (a)	**14.** (c)	**15.** (b)	**16.** (b)	**17.** (d)	**18.** (b)
19. (b)	**20.** (a)	**21.** (a)	**22.** (d)	**23.** (b)	**24.** (a)	**25.** (c)		

Review Self-practice Problems

15.20 (a) Given $r_{12} = 0.5$, $r_{13} = 0.4$, and $r_{23} = 0.1$, find the values of $r_{12.3}$ and $r_{23.1}$.
[*Kerala Univ., M.Com 2007*]

(b) If $r_{12} = 0.60$, $r_{13} = 0.70$, $r_{23} = 0.65$, and $S_1 = 1.0$, find $S_{1.23}$, $R_{1.23}$, and $r_{12.3}$.
[*Kurukshetra Univ., M.Com, 2008*]

15.21 (a) If $r_{12} = 0.80$, $r_{13} = -0.56$ and $r_{23} = 0.40$, then obtain $r_{12.3}$ and $R_{1.23}$.
[*Saurashtra Univ., M.Com, 2007*]

(b) In a three variables distribution, it is found that $r_{13} = 0.6$, $r_{23} = 0.5$, and $r_{12} = 0.8$. Find the value of $r_{13.2}$.
[*Madras Univ., M.Com, 2006*]

15.22 In a three variables distribution

$\bar{x}_1 = 28.20$ $\bar{x}_2 = 4.91$ $\bar{x}_3 = 594$
$s_1 = 4.4$ $s_2 = 1.1$ $s_3 = 80$
$r_{12} = 0.80$ $r_{23} = -0.56$ $r_{31} = -0.40$

(a) Find the correlation coefficient $r_{23.1}$ and $R_{1.23}$.

(b) Also estimate the value of x_1 when $x_2 = 6.0$ and $x_3 = 650$.

15.23 The following data relate to agricultural production (x_1) in quintal/hectare, rainfall (x_2) in inches, and use of fertilizers (x_3) in kg/hectare.

x_1 :	85	76	82	83	72	93	76	81
x_2 :	6	8	14	11	9	16	5	3
x_3 :	40	25	5	20	15	10	35	50

Find $R_{1.23}$ and the coefficient of a multiple determination and interpret your result.

15.24 In a study of the factors: honors points (x_1), general intelligence (x_2), and hours of study (x_3), which influence 'academic success', a statistician obtained the following results based on the records of 450 students at a university campus.

$\bar{x}_1 = 18.5, \quad \bar{x}_2 = 100, \quad \bar{x}_3 = 24$

$s_1 = 11.2, \quad s_2 = 15.8, \quad s_3 = 6$

$r_{12} = 0.60, \quad r_{13} = 0.32, \quad r_{23} = -0.35$

Find to what extent honors points are related to general intelligence when hours of study per week are held constant. Also find the other partial correlation coefficients.

15.25 In a three variables distribution, the following data have been obtained

x_1 :	3	5	6	8	12	14
x_2 :	16	10	7	4	3	2
x_3 :	90	72	54	42	30	12

(a) Find a regression equation of x_3 on x_1 and x_2

(b) Estimate the value of x_3 for $x_1 = 10$ and $x_2 = 6$

(c) Find the standard error of estimate of x_3 on x_1 and x_2

(d) Determine a multiple correlation coefficient $R_{3.12}$

Hints and Answers

15.20 (a) $r_{12.3} = 0.504$; $r_{23.1} = -0.126$

(b) $S_{1.23} = 0.688$; $R_{1.23} = 0.726$; $r_{12.3} = 0.267$

15.21 (a) $r_{12.3} = 0.759$; $R_{1.23} = 0.842$

(b) $r_{13.2} = 0.385$

15.22 (a) $r_{23.1} = -0.436$; $R_{1.23} = 0.802$

(b) Regression equation of x_1 on x_2 and x_3

$$x_1 - \bar{x}_1 = \left(\frac{r_{12} - r_{13}\, r_{23}}{1 - r_{23}^2}\right)\left(\frac{s_1}{s_2}\right)(x_2 - \bar{x}_2) + \left(\frac{r_{13} - r_{12}\, r_{23}}{1 - r_{23}^2}\right)\left(\frac{s_1}{s_3}\right)(x_3 - \bar{x}_3)$$

$$x_1 - 28.02 = \left\{\frac{0.8 - (-0.4)(-0.56)}{1 - (-0.56)^2}\right\}\left(\frac{4.4}{1.1}\right)(x_2 - 4.91) + \left\{\frac{(-0.4) - (0.8)(-0.56)}{1 - (-0.56)^2}\right\}\left(\frac{4.4}{80}\right)(x_3 - 594)$$

$$x_1 = 9.225 + 3.356x_2 + 0.003x_3$$

For $x_2 = 6$ and $x_3 = 650$, we have $x_1 = 31.896$.

15.23 $R_{1.23} = 0.975$; $R_{21.23} = 0.9512$ implies that 95.12 per cent variation in agriculture production is explained.

15.24 $r_{12.3} = 0.80$; $r_{13.2} = 0.71$ and $r_{23.1} = -0.721$

15.25 (a) $x_3 = 61.40 - 3.65x_1 + 2.54x_2$

(b) Estimated value of $x_3 = 40$

(c) $S_{3.12} = 3.12$ (d) $R_{3.12} = 0.9927$