UNIT 17 DISTRIBUTIONS OF CORRELATION COEFFICIENTS

Structure **Page No**

17.1 INTRODUCTION

We shall begin this unit by discussing the distribution of correlation coefficients in Sec. 17.2. As we have already studied the correlation coefficient of two variables, in this unit, we shall extend this to more than two variables by introducing the multiple correlation and partial correlation coefficients. The distribution of partial correlation coefficient and distribution of multiple correlation coefficients will be discussed in Sec. 17.4 and Sec. 17.5, respectively. At the end, we shall focus on the distribution of $\frac{R^2}{1-R^2}$ in Sec. 17.5.

Objectives

After studying this unit, you should be able to:

- define the correlation coefficients;
- calculate the correlation coefficient for multiple and partial correlation;
- describe the distribution of multiple correlation coefficients;
- describe the distribution of partial correlation coefficients;
- find the distribution of $R^2/(1-R^2)$.

17.2 DISTRIBUTION OF CORRELATION COEFFICIENTS

Let us start with the linear regression analysis with two independent variables and one dependent variable. This is nothing but the extention of two variable regression analyses. Let us assume the independent variables be x_1 and x_2 and the dependent variable be y. Suppose we experimentally take observations on y by setting (x_1, x_2) at n different levels. Let y_i be the observed value of y when (x_1, x_2) is set of (x_{1i}, x_{2i}) where $i = 1, 2, \ldots, n$. Since we are assuming a linear relationship, the following model is considered $y_i = b_0 + b_1x_{1i} + b_2x_{2i}$, where $i = 1, 2, \ldots, n$.

In this model, because of random variation in y, the points (y_i, x_{1i}, x_{2i}) may not lie on a plane, thus an error term e_i is added to the equation of plane. Next we estimate b_0, b_1, b_2 using observations y_i, so that we get a best fit linear model.

Now let $\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i}$, $i = 1, 2, \ldots, n$ (1)

where $\hat{b}_0, \hat{b}_1$ and $\hat{b}_2$ are the estimated parameters and then $\hat{y}_i$ is the best estimated value of y_i. The best fit plane is known as regression plane.

Let us minimize the sum of squares of $y_i - \hat{y}_i$ the residuals or errors, that is

$$\sum_{i=1}^{n} e^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$= \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{1i} - b_2 x_{2i})^2 \text{ (Using Eqn. (1))} \quad (2)$$

For minimization, the necessary condition is that the partial derivatives of expression in Eqn. (2) with respect to b_0, b_1 and b_2 are zero. Thus

$$-\sum_{i=1}^{n} (y_i - b_0 - b_1 x_{1i} - b_2 x_{2i}) = 0$$

$$-\sum_{i=1}^{n} x_{1i} (y_i - b_0 - b_1 x_{1i} - b_2 x_{2i}) = 0 \quad (3)$$

$$-\sum_{i=1}^{n} x_{2i} (y_i - b_0 - b_1 x_{1i} - b_2 x_{2i}) = 0$$

These equations can be written as

$$\sum y_i = b_0 n + b_1 \sum x_{1i} + b_2 \sum x_{2i} \quad (4)$$

$$\sum x_{1i} y_i = b_0 \sum x_{1i} + b_1 \sum x_{1i}^2 + b_2 \sum x_{1i} x_{2i} \quad (5)$$

$$\sum_{i=1}^{n} x_{2i} y_i = b_0 \sum x_{2i} + b_1 \sum x_{1i} x_{2i} + b_2 \sum x_{2i}^2 \quad (6)$$

The system of equations in Eqns. (4), (5) and (6) are the normal equations of the least square method. This system of normal equations can be solved to find the values of b_0, b_1 and b_2 defined by $\hat{b}_0, \hat{b}_1$ and $\hat{b}_2$ and can be substituted in Eqn. (1) to get the equation of regression plane. The coefficients of independent variables b_1 and b_2 are regression coefficients for linear regression. This model can be generalized to handle the prediction a dependent variable forms several independent variables. Let us consider the independent variables $x_1, x_2, \ldots, x_k$, and the dependent variable y. Then linear model is

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik} + e_i; \; i = 1, 2, 3, \ldots, n \quad (7)$$

where e_i is the error or the difference between the true value and the estimated value. The matrix notation of Eqn. (7) is

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ 1 & \vdots & & & \\ \vdots & & & & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

which can be rewritten as $\mathbf{y} = \mathbf{x}\mathbf{b} + \mathbf{e}$ (8)

The errors e_i are random variables for which the following assumptions are made

i) $E(e_i) = 0$

ii) $\operatorname{cov}(e_i, e_j) = 0, \; i \neq j.$

iii) $\operatorname{var}(e_i) =$ constant $(\text{say } \sigma^2)$.

Now, let us extend the least square method to more than two independent variables. For this, we rewrite Eqn. (7) as given below

$$e_i = y_1 - b_0 + b_1x_{i1} - b_2x_{i2} \cdots b_kx_{ik}$$

To fit the model, we estimate the values for the regression coefficients $b_0, b_1, \ldots, b_k$ with the available data. For this, according to the method of least square, we minimize the sum of squares of the errors w.r.t b_i's. Thus, we minimize

$$\mathbf{e'e} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - b_0x_{i1} - b_1x_{i2} \cdots b_kx_{ik}\right)^2$$

$$= (\mathbf{y} - \mathbf{xb})'(\mathbf{y} - \mathbf{xb}) \qquad \text{[using Eqn. (8)]} \qquad (9)$$

Now we get a system of k + 1 normal equations (similar to Eqn. (4), (5), (6)). If we assume that rank of matrix x in Eqn. (8) is (k + 1), then normal equations will have a unique solution given by Eqn.(11).

Let us denote the estimate or fitted value of b by $\hat{b}$. Then the deviation of the estimated value and the true value is

$$\hat{e}_i = \hat{y}_i - \hat{b}_0 - \hat{b}_1x_{i1} + \hat{b}_kx_{ik};\ i = 1, 2, \ldots, n \qquad (10)$$

or $\quad \hat{\mathbf{e}} = \mathbf{y} - \mathbf{x}\hat{\mathbf{b}}$

where $\hat{\mathbf{e}}$ is the vector of estimated residuals $\hat{e}_i$.

The least square estimate of b, obtained by solving normal equations, is given by

$$\hat{\mathbf{b}} = (\mathbf{x'x})^{-1}\mathbf{x'y} \qquad (11)$$

As noted above the rank of x should be $k + 1 \le n$.

Using the value of $\hat{b}$ given in Eqn. (11) and Eqn.(8), the fitted value of y can be written as

$$\hat{\mathbf{y}} = \mathbf{x}\hat{\mathbf{b}} = \mathbf{x}(\mathbf{x'x})^{-1}\mathbf{x'y} \qquad (12)$$

The matrix $\mathbf{x}(\mathbf{x'x})^{-1}\mathbf{x'}$ is known as 'Hat matrix'. The residual vector is

$$\hat{e} = y - \hat{y} \qquad (13)$$

$$= \left[I - x(x'x)^{-1}x'\right]y \qquad \text{[using Eqn. (12)]}$$

and the residual sum of squares equal $\hat{\mathbf{e}}'\mathbf{e}$.

Now let us illustrate this concept in the following examples.

Example 1: Formulate a linear regression model in matrix form for the following data:

x_1	0	1	2	3	4
x_2	10	7	5	4	3
y	1	4	3	8	9

Solution: The independent variables are x_1 and x_2 and the dependent variable is y. The number of observations given here is 5. The corresponding matrices are

$$y = \begin{bmatrix} 1 \\ 4 \\ 3 \\ 8 \\ 9 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} 1 & 0 & 10 \\ 1 & 1 & 7 \\ 1 & 2 & 5 \\ 1 & 3 & 4 \\ 1 & 4 & 3 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \qquad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix}$$

Therefore, we get $\mathbf{y} = \mathbf{xb} + \mathbf{e}$

Example 2: The owner of a chain of five stores wishes to forecast net profit with the help of next year's projected sales of food and non-food items. The data about current year's sales of food items, sales of non-food items as also net profit for all the five stores are available as follows:

Table 1: Sales of Food and Non-Food Item and Net Profit of a Chain of Stores

Store No.	Net Profit (Rs. Crores) y	Sales of Food Items (Rs. Crores) x_1	Sales of Non-Food Items (Rs. Crores) x_2
1	15	5	2
2	9	3	3
3	3	4	3
4	7	2	6
5	3	1	7

Assuming a linear regression model $y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + e_i$, where $i = 1, 2, 3, 4, 5$, calculate

i) the least square estimates $\hat{\mathbf{b}}$,
ii) the residuals $\hat{\mathbf{e}}$ and
iii) the residual sum of squares for the model.

Solution: (i) From the given data,

$$\mathbf{x} = \begin{bmatrix} 1 & 5 & 2 \\ 1 & 3 & 3 \\ 1 & 4 & 3 \\ 1 & 2 & 6 \\ 1 & 1 & 7 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 15 \\ 9 \\ 3 \\ 7 \\ 3 \end{bmatrix} \text{ and Rank } \mathbf{x} = 3$$

$$\mathbf{x}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 5 & 3 & 4 & 2 & 1 \\ 2 & 3 & 3 & 6 & 7 \end{bmatrix} \quad \mathbf{x}'\mathbf{x} = \begin{bmatrix} 5 & 15 & 21 \\ 15 & 55 & 50 \\ 21 & 50 & 107 \end{bmatrix}$$

$$(\mathbf{x}'\mathbf{x})^{-1} = \begin{bmatrix} \frac{677}{19} & \frac{-111}{19} & \frac{-81}{19} \\ \frac{-111}{19} & \frac{94}{95} & \frac{13}{19} \\ \frac{-81}{19} & \frac{13}{19} & \frac{10}{19} \end{bmatrix} \text{ and } \mathbf{x}'\mathbf{y} = \begin{bmatrix} 37 \\ 131 \\ 129 \end{bmatrix}$$

$$\hat{\mathbf{b}} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}$$

$$= \begin{bmatrix} 59/19 \\ 164/95 \\ -4/19 \end{bmatrix},$$

The fitted regression plane is $\hat{y}_i = \frac{59}{19} + \frac{164}{95} x_{1i} - \frac{4}{19} x_{2i}$, $i = 1, 2, \ldots, 5$.

(ii) $\hat{y} = x\hat{b} = \begin{bmatrix} 215/19 \\ 727/95 \\ 891/95 \\ 503/95 \\ 319/95 \end{bmatrix}$

and the residuals $\hat{e} = y - \hat{y} = \begin{bmatrix} 70/19 \\ 128/95 \\ -606/95 \\ 162/95 \\ -34/95 \end{bmatrix}$

(iii) The residual sum of squares is

$$\hat{e}'\hat{e} = \frac{5619}{95}.$$

Now, try the following exercises.

E1) Given the following data, fit a regression equation representing dependence of number of credit cards on family size and family income. Also, show whether addition of 'Family Income' variable has improved the relationship by finding sums of squares of errors as also calculating simple and multiple correlation coefficients.

No. of Credit Cards y	Family Size x_1	Family Income (Rs. In lakhs) x_2
2	0	2
3	2	6
2	2	7
7	2	5
6	4	9
8	4	8
10	4	7
7	6	10
8	6	11
12	6	9
11	8	15
14	8	13

E2) If the Hat matrix, given in Eqn. (12), is $H = x(x'x)^{-1}x'$, then verify that

i) $\hat{e} = I - H$ is symmetric.

ii) $\hat{e} = I - H$ is idempotent.

So far, we have discussed the linear regression fit. Now, we shall obtain some kind of coefficient of correlation that measures the adequacy of simultaneous linear fit with several independent variables. For this consider $y'y$.

$$y'y = (\hat{y} + \hat{e})'(\hat{y} + \hat{e}) \qquad \text{[using Eqn. (13)]}$$

$$y'y = \hat{y}'\hat{y} + \hat{e}'\hat{y} + \hat{y}'\hat{e} + \hat{e}'\hat{e}$$

$$y'y = \hat{y}'\hat{y} + \hat{e}'\hat{e} \qquad \text{[since } \hat{e}'\hat{y} = 0\text{, using Eqn.(13)]}$$

$$y'y - n\bar{y}^2 = \hat{y}'\hat{y} - n\left(\bar{\hat{y}}\right)^2 + \hat{e}'\hat{e} \quad \text{[since } \bar{y} = \bar{\hat{y}} \text{ by normal Eqn. (4) for general k]}$$

where $\bar{y}$ = mean of y_i series

$\bar{\hat{y}}$ = mean of $\hat{y}_i$ series.

or $$\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = \sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2 + \sum_{i=1}^{n}\hat{e}_i^2$$

or Total sum of squares = Regression sum of squares + Residual (or error) sum of squares

Dividing both the sides by $\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$ and rearranging the terms, we get

$$\frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} = 1 - \frac{\sum_{i=1}^{n}\hat{e}_i^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$$

The sum of squares breakup suggests the following measure for goodness of fit for linear model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\hat{e}_i^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{14}$$

R^2 is the coefficient of multiple determination. It can be shown that R^2 is the square of Pearson product moment correlation coefficient between the observed values y_i and the fitted values $\hat{y}_i$. This correlation coefficient is called multiple correlation coefficient between y and $x_1, x_2, \ldots, x_k$

Thus, we see that

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y} - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{\text{Regression sum of squares}}{\text{Total sum of squares about mean}}$$

where R^2 is the squared multiple correlation coefficient. We write R^2 with the subscripts as corresponding variables under consideration. For example, if we assume the dependent variable y and the independent variables x_1, x_2, x_3, then the squared multiple correlation coefficient is denoted as $R^2_{y \cdot x_1 x_2 x_3}$. The value of R^2 lies between 0 to 1. If the value of $R^2 = 1$, then the fitted equation passes through each of data points and in that case $\hat{e}_i = 0$ for all i. While $R^2 = 0$ implies $\hat{y}_i = \bar{y}$ for all i or the values of $x_i's$ denote affect the estimates of y_i.

Now, in the following section, let us discuss the multiple correlation coefficient in the case when independent variables are not considered as fixed as in the present section but are random variables.

17.3 MULTIPLE CORRELATION COEFFICIENTS

If the modeling is done by fitting a linear equation among more than two variables i.e. more than one independent variables, then the extent of fit is measured by correlation coefficient of dependent variable and independent variables called the multiple correlation coefficient.

Let us consider the set of independent variables or predictor variables $x_1, x_2, x_3, \ldots, x_k$, and the dependent or response variable y. In many applications $(y, x_1, \ldots, x_k)$ be y_i for $i = 1, 2, \ldots, n$, can be considered random vector having some joint probability distribution and the problem is to find a best linear function of $x_1, x_2, \ldots, x_k$, say, $b_0 + b_1x_2 + \cdots + b_kx_k$ to estimate or predict y.

Let $\quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}, \ \mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix}$

Then y is predicted by $b_0 + \mathbf{b'x}$. Then $e = b_0 - \mathbf{b'x}$ is the prediction error.

Let $\mathbf{z} = \begin{pmatrix} y \\ \mathbf{x} \end{pmatrix}$ and $\Sigma = \begin{bmatrix} \underset{1\times 1}{\sigma_{yy}} & \vdots & \underset{1\times k}{\boldsymbol{\sigma}'_{yx}} \\ \cdots & \vdots & \cdots \\ \underset{k\times 1}{\boldsymbol{\sigma}_{yx}} & \vdots & \underset{k\times k}{\boldsymbol{\sigma}_{xx}} \end{bmatrix}$ be its corresponding partitioned variance covariance matrix. Σ is assumed to be positive definite. The multiple correlation coefficient between y and the predictor variables $x_1, x_2, \ldots, x_k$ is defined to be the maximum correlation coefficient between y and any linear function $b'x = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k$. The multiple correlation coefficient $R_{y.x_1x_2\cdots x_k}$ is thus defined by

$$R_{y.x_1x_2\cdots x_k} = R_{y.x} = \overset{\max.}{b} \frac{\operatorname{cov}(y, b'x)}{\sqrt{\operatorname{var}(y)\operatorname{var}(b'x)}} \tag{15}$$

$$= \overset{\max}{b} \frac{\mathbf{b}'\boldsymbol{\sigma}_{yx}}{\sqrt{\sigma_{yy}\mathbf{b}'\Sigma_{xx}\mathbf{b}}} \quad \text{[variance or covariance does not depend on } b_0]$$

$$= \overset{\max}{b} \frac{\mathbf{b}'\Sigma_{xx}^{1/2}\Sigma_{xx}^{-1/2}\sigma_{yx}}{\sqrt{\sigma_{yy}\mathbf{b}'\Sigma_{xx}\mathbf{b}}}$$

$$= \overset{\max}{b} \frac{\left(\Sigma_{xx}^{1/2}\mathbf{b}\right)'\left(\Sigma_{xx}^{-1/2}\sigma_{yx}\right)}{\sqrt{\sigma_{yy}b'\Sigma_{xx}b}}.$$

Further $\dfrac{\left(\Sigma_{xx}^{1/2}\mathbf{b}\right)'\left(\Sigma_{xx}^{1/2}\sigma_{yx}\right)}{\sqrt{\sigma_{yy}\mathbf{b}'\Sigma_{xx}\mathbf{b}}} \le \dfrac{\sqrt{\left(\Sigma_{xx}^{1/2}\mathbf{b}\right)'\left(\Sigma_{xx}^{1/2}\mathbf{b}\right)}\sqrt{\left(\Sigma_{xx}^{-\frac{1}{2}}\sigma_{yx}\right)'\left(\Sigma_{xx}^{-1/2}\sigma_{yx}\right)}}{\sqrt{\sigma_{yy}\mathbf{b}'\Sigma_{xx}\mathbf{b}}}$

for all b_0, **b**. [using Cauchy-Schwarz inequality $u'v \le \sqrt{u'u}\sqrt{v'v}$]

$$= \frac{\sqrt{\left(\Sigma_{xx}^{-1/2}\sigma_{yx}\right)'\left(\Sigma_{xx}^{-1/2}\sigma_{yx}\right)}}{\sqrt{\sigma_{yy}}}$$

$$= \frac{\sqrt{\sigma'_{yx}\Sigma_{xx}^{-1}\sigma_{yx}}}{\sqrt{\sigma_{yy}}}$$

Clearly the equality is attained when $\mathbf{b} = \Sigma_{xx}^{-1}\sigma_{yx}$

Therefore,

$$R_{y.x}^2 = \frac{\sigma'_{yx}\Sigma_{xx}^{-1}\sigma_{yx}}{\sigma_{yy}} \tag{16}$$

and $R_{y.x}$ is the correlation between y and its best linear predictor, and is known as the population multiple correlation coefficient. The square of R that is R^2 is called the population coefficient of determination.

It may be noted that $0 \le R \le 1$ by definition. If $R = 0$, then x has no predictive power and if $R = 1$, then y can be predicted without error. In the following we assume z has multivariate normal distribution $N_k(\mu, \Sigma)$. It may be recalled that if $Z \sim N_{k+1}(\mu, \Sigma)$, then the conditional distribution of y given x is normal with mean

$$E(y/x) = \mu_y + \sigma'_{yx}\sum\nolimits_{xx}^{-1}(X - \mu_x) \tag{17}$$

and variance

$$\text{var}(y/x) = \sigma_{yy} - \sigma'_{yx}\sum\nolimits_{xx}^{-1}\sigma_{yx} \tag{18}$$

Thus in present case best linear predictor of y is in fact the regression of y on x and

$$\text{Var}\left(y/x\right) = \sigma_{yy} - \sigma_{yy}R_{y.x}^2 = \sigma_{yy}\left(1 - R_{y.x}^2\right) \tag{19}$$

Even if z is not multivariate normal but the linear model
$y = b_0 + b_1x_1 + \cdots + b_kx_k + e = b_0 + \mathbf{b}'x + e$ is fitted so that the Mean squared error $= E(y - b_0 - b'x)^2$ is minimal. Then it can be proved that minimum mean squared error equals

$$\text{var}(y/x) = \sigma_{yy} - \sigma'_{yx}\sum\nolimits_{xx}^{-1}\sigma_{yx}. \qquad \text{[as in Eqn. (18)]}$$

and is attained when $\mathbf{b} = \Sigma_{xx}^{-1}\boldsymbol{\sigma}_{yx}$.

Now let us understand this in the following examples.

Example 3: Derive the coefficient of correlation of bivariate case from the multiple correlation coefficient.

Solution: For bivariate case, let the predictor variable be x and the response variable be y. Then, we have

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x & \sigma_y \\ \rho\sigma_x & \sigma_y & \sigma_y^2 \end{bmatrix}$$

where σ_x and σ_y be variances of x and y and the correlation coefficient between them be ρ.

Then $\text{var}(y/x) = \sigma_y^2(1 - \rho^2)$,

$$R_{xy}^2 = \frac{\sigma_y^2 - \sigma_y^2(1 - \rho^2)}{\sigma_y^2}$$

$$= \rho^2$$

and hence

$$R_{xy} = |\rho|.$$

Example 4: Consider the mean vector be $\mu_x = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$ and $\mu_y = 4$, and the covariance matrices of x_1, x_2 and y are $\sum_{xx} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \sigma_{yy} = 9, \sigma_{xy} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

Evaluate

i) Fit the equation $y = b_0 + b_1x_1 + b_2x_2$ as best linear equation,

ii) The multiple correlation coefficient, and

iii) The mean square error.

Solution: i) $\mathbf{b} = \sum_{xx}^{-1} \sigma_{xy}$

$$= \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

$$b_0 = \mu_y - b'\mu_x = 4 - \begin{bmatrix} -1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = 11$$

Therefore,

$$y = b_0 + b'x = 11 + 2\ x_1 - x_2.$$

ii) $$R^2 = \frac{\sigma_{xy} \cdot \sum_{xx}^{-1} \sigma_{xy}}{\sigma_{yy}}$$ [using Eqn. (16)]

$$= \frac{\begin{bmatrix} 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ 1 \end{bmatrix}}{9}$$

$$= \frac{1}{9} \left[\begin{bmatrix} 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right]$$

$$= \frac{1}{9} \left[\begin{bmatrix} 3 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right]$$

$$= \frac{5}{9}$$

Or $$R = \frac{\sqrt{5}}{3} = 0.75.$$

iii) The mean square error $= \sigma_{yy}\left(1 - R^2\right)$ [using Eqn. (19)]

$$= 9\left(1 - \frac{5}{9}\right)$$

$$= 4.$$

Now, try the following exercises.

E3) The following data gives DPS (dividend per share), EPS (Earning per share) and Return on Market, for six companies.

DPS (y)	EPS (x_1)	Return on Market (x_2)
12	8	1
6	3	2
3	5	2
22	17	5
4	9	6
10	16	8

Find the regression equation $y = b_0 + b_1x_1 + b_2x_2$. Find the multiple correlation coefficient and mean square error.

So far, we have discussed the population multiple correlation coefficient. Under the assumption of normality of z, let us now discuss the estimation of population multiple correlation coefficient. The estimate is called sample multiple correlation coefficient. For this consider the joint distribution of y with $x_1, x_2, ..., x_k$ is $N_{k+1}(\mu, \Sigma)$ where

$$\begin{bmatrix} \mu_y \\ .. \\ \mu_x \end{bmatrix} \text{ and } \Sigma = \begin{pmatrix} \sigma_{yy} & \vdots & \sigma'_{yx} \\ \cdots & \vdots & \cdots \\ \sigma_{yx} & \vdots & \sigma_{xx} \end{pmatrix}.$$

Now suppose we have sample of size n on (y, x_1,...,x_k) and we computes the sample mean vector $\hat{\mu}$ and sample covariance matrix S for this random sample. Let these be partitioned as

$$\hat{\mu} = \begin{bmatrix} \bar{y} \\ \cdots \\ \bar{x} \end{bmatrix} \text{ and } S = \begin{bmatrix} s_{yy} & \vdots & s'_{yx} \\ \cdots & \cdots & \cdots \\ s_{yx} & \vdots & S_{xx} \end{bmatrix} \qquad (20)$$

It can be shown that the maximum likelihood estimators of the coefficients $b_0, \mathbf{b}$ are

$$\hat{\mathbf{b}} = S_{xx}^{-1} s_{yx} \text{ and} \qquad (21)$$

$$b_0 = \bar{y} - \hat{\mathbf{b}}'\bar{\mathbf{x}} = \bar{y} - s'_{yx}S_{xx}^{-1}\bar{x}. \qquad \text{[using Eqn. (21)]} \qquad (22)$$

Now $\quad b_0 + \hat{\mathbf{b}}'\mathbf{x} = \bar{y} + s'_{xy}S_{xx}^{-1}(\mathbf{x} - \bar{x}) \qquad (23)$

It may be noted that $b_0 + \hat{b}'x$ is called the linear regression function. The maximum likelihood estimator of $R_{y.x}$ is the sample multiple correlation coefficient which is defined as

$$\hat{R}_{y.x} = \sqrt{\frac{s'_{yx}S_{xx}^{-1}s_{yx}}{s_{yy}}}. \qquad (24)$$

If the considered distribution is normal, then $\hat{R}$ is the maximum likelihood estimate of R.

Now, in the following section let us discuss the partial correlation coefficient.

17.4 PARTIAL CORRELATION COEFFICIENTS

Let us consider the variables y, x_1 and x_2 where y is the linear predictor of two independent variables x_1 and x_2. If we consider the simple correlation coefficient between y and x_1 or between y and x_2 keeping x_2 or x_1 constant respectively, then this correlation coefficient is said to be partial correlation coefficient. Therefore, the partial correlation coefficient is determined in terms of simple correlation coefficients among the various variables involved in a multiple relationship. In the case of three variables y, x_1 and x_2 there are two partial correlation coefficients, which are

$r_{yx_2 \cdot x_3}$ = partial correlation coefficient between y and x_2 keeping x_3 constant.

$r_{yx_3 \cdot x_2}$ = partial correlation coefficient between y and x_3 keeping x_2 constant.

Expanding the number of variables, the partial correlation coefficient between y_i and x_i holding $x_{p+1}, x_{p+2}, \ldots, x_k$ fixed is given by

$$r_{ij \cdot p+1 \cdots k} = \frac{\sigma_{ij \cdot p+1 \cdots k}}{\sqrt{\sigma_{ij \cdot p+1 \cdots k}}\sqrt{\sigma_{ij \cdot p+1 \cdots k}}} \tag{25}$$

Let us define the distribution of this by dividing the components y and x into two groups $X^{(1)}$ and $X^{(2)}$. Accordingly, the corresponding mean vectors μ of X is divided into $\mu^{(1)}$ and $\mu^{(2)}$ and covariance matrix Σ is divided into $\Sigma_{11}, \Sigma_{12}, \Sigma_{22}$, where Σ_{11}, Σ_{12} and Σ_{22} are the covariance matrices of $X^{(1)}$, of $X^{(1)}$ and $X^{(2)}$, and of $X^{(2)}$, respectively. Further assuming that the distribution is normal, then the joint distribution of $X^{(1)}$ given $X^{(2)}$, is normal with mean $\mu^{(1)} + \Sigma_{12}, \Sigma_{22}^{-1}\left(X^{(2)} - \mu^{(2)}\right)$ and the covariance matrix is $\Sigma_{11 \cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Example 5: Consider the above in a bivariate normal distribution to find the conditional distribution of $X^{(1)} = X_1$ given $X^{(2)} = X_2 = x_2$.

Solution: Given is $X^{(1)} = X_1$ and $X^{(2)} = X_2$,

therefore $\mu^{(1)} = \mu_1$ and $\mu^{(2)} = \mu_2$,

$$\Sigma_{11} = \sigma_1^2, \Sigma_{12} = \Sigma_{21} = \rho\sigma_1\sigma_2 \text{ and } \Sigma_{22} = \sigma_2^2$$

Therefore

$$\Sigma_{11 \cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$$= \sigma_1^2 - \frac{\rho^2\sigma_1^2\sigma_2^2}{\sigma_2^2} = \sigma_1^2\left(1-\rho^2\right)$$

Example 6: Consider the random vector **z** be

$$N_4(\mu, \Sigma) \text{ where } \mu = \begin{bmatrix} 2 \\ 5 \\ -2 \\ 1 \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} 9 & 0 & 3 & 3 \\ 0 & 1 & -1 & 2 \\ 3 & -1 & 6 & -3 \\ 3 & 2 & -3 & 7 \end{bmatrix}$$

If z is partitioned as $z = \left(Y_1, Y_2, X_1, X_2\right)'$ and $\mu_y = \begin{pmatrix} 2 \\ 5 \end{pmatrix}, \mu_x = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$, $\Sigma_{yy} = \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix}$

$\Sigma_{yx} = \begin{bmatrix} 3 & 3 \\ -1 & 2 \end{bmatrix}$ and $\Sigma_{xx}' = \begin{bmatrix} 6 & -3 \\ -3 & 7 \end{bmatrix}$, Then find

$E(Y/X)$, $\text{cov}(Y/X)$, r_{12}, $r_{12\cdot34}$ and $r_{13\cdot24}$.

Solution: $E(Y/X) = \mu_y + \Sigma_{yx}\ \Sigma_{xx}^{-1}\left(X - \mu_x\right)$

$$= \begin{bmatrix} 2 \\ 5 \end{bmatrix} + \begin{bmatrix} 3 & 3 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 6 & -3 \\ -3 & 7 \end{bmatrix}^{-1}\begin{bmatrix} x_1 + 1 \\ x_2 - 1 \end{bmatrix}$$

$$= \begin{bmatrix} 3+\frac{10}{11}x_1+\frac{9}{11}x_2 \\ \frac{14}{3}-\frac{1}{33}x_1+\frac{3}{11}x_2 \end{bmatrix}$$

$$\text{cov.}(Y/X) = \Sigma_{yy} - \Sigma_{yx}\,\Sigma_{xx}^{-1}\Sigma_{xy}$$

$$= \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 3 & 3 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ -3 & 7 \end{bmatrix}^{-1} \begin{bmatrix} 3 & -1 \\ 3 & 2 \end{bmatrix}$$

$$= \frac{1}{33}\begin{bmatrix} 126 & -24 \\ -24 & 14 \end{bmatrix}$$

$$r_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} = \frac{0}{\sqrt{9\times 1}} = 0$$

$$r_{12\cdot 34} = \frac{\sigma_{12\cdot 34}}{\sqrt{\sigma_{12\cdot 34}\sigma_{22\cdot 34}}} = \frac{-24/33}{\sqrt{\frac{126}{33}\cdot\frac{14}{33}}} = \frac{-24}{\sqrt{126\cdot 14}}$$

$$= -\frac{4}{7} = -0\cdot 571.$$

Now, try the following exercise.

E4) Let y be distributed as $N_3(\mu, \Sigma)$ where $\mu = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{bmatrix}$

Find

i) the distribution of $z = 4y_1 - 6y_2 + y_3$ and $z = \begin{bmatrix} y_1 - y_2 + y_3 \\ 2y_1 + y_2 - y_3 \end{bmatrix}$

ii) conditional densities $f(y_2 | y_1, y_3)$ and $f(y_1, y_2 | y_3)$

iii) r_{12} and $r_{12\cdot 3}$.

S_{yy}, S_{yx}, S_{xx} are the estimators of $\Sigma_{yy}, \Sigma_{yx}, \Sigma_{xx}$. Therefore, the estimator of $r_{ij\cdot rr\cdots q}$ is $\hat{r}_{ij\cdot rr\cdots q}$, which is the ij[th] element of $R_{y\cdot x} = D_s^{-1}\left(S_{yy} - S_{yx}S_{xx}^{-1}S_{xy}\right)D_s^{-1}$ where $D_s = \left[\text{diag}\left(S_{yy} - S_{yx}S_{xx}^{-1}S_{xy}\right)\right]^{1/2}$.

Now let us define the sample partial correlation coefficient between y_1 and y_2 with y_3 fixed as

$$r_{12\cdot 3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$$

where r_{12}, r_{13} and r_{23} are the ordinary correlation coefficients between y_1 and y_2, y_1 and y_3 and y_2 and y_3 respectively.

In the next section, we shall discuss distribution of $R^2/(1-R^2)$.

17.5 DISTRIBUTION OF $\frac{R^2}{1-R^2}$

Let us first recall the multiple correlation coefficient given in Eqn. (14). This is an alternative way of computing the multiple coefficient of correlation. Before we proceed for the distribution, let us define and discuss the following theorem.

Theorem 1: let X be a random matrix of order $n \times k$ and y be a random matrix of order $n \times 1$ with $P(y=0)=0$. Suppose x is independent of y and the rank of X is $k-1$ with probability 1. If the sample multiple correlation coefficient $\hat{R}$ given in Eqn. (24), then $\hat{R}^2$ has the beta distribution with parameters $\frac{1}{2}(k-1)$ and $\frac{1}{2}(n-1)$, which can also be written as $\frac{n-k}{k-1}.\frac{\hat{R}^2}{1-\hat{R}^2}$ is $F_{k-1,\, n-k}$.

Proof: Consider the partitions given as

$$\mu = \begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix} \text{ and } \Sigma = \begin{bmatrix} \sigma_{yy} & 0' \\ 0 & \Sigma_{xx} \end{bmatrix}$$

where the order μ_x is $(k-1)\times 1$ and order of Σ_{xx} is $(k-1)\times(k-1)$. Here, it may be noted that since $R=0$, therefore $\sigma_{12}=0$. Here the multiple correlation between $\frac{y_i-\mu_y}{\sqrt{\sigma_{yy}}}$ and $\sum_{xx}^{-1/2}(X_i-\mu_x)$ is same as that of between Y_i and X_1, therefore it can be assumed that $\mu=0$ and $\Sigma=I_k$. Now all the conditions of Theorem 1 are satisfied, and the result follows.

Example 7: Find the r^{th} moment of R^2 and hence derive mean and variance of $\hat{R}^2$.

Solution: The r^{th} moment of $\hat{R}^2$ is

$$E\left(\hat{R}^{2r}\right) = \frac{\gamma\left[\frac{1}{2}(n-1)\right]\gamma\left(\frac{1}{2}(k-1)+x\right)}{\gamma\left(\frac{1}{2}(n-1)+x\right)\gamma\left(\frac{1}{2}(k-1)\right)}$$

$$= \frac{\left[\frac{1}{2}(k-1)\right]r}{\left[\frac{1}{2}(n-1)\right]r}$$

The mean $= E\left(\hat{R}^2\right) = \frac{k-1}{n-1}$

and the variance is $E\left(\hat{R}^4\right) = \frac{k^2-1}{n^2-1} - \left(\frac{k-1}{n-1}\right)^2$

$$= \frac{2(n-k)(k-1)}{(n^2-1)(n-1)}.$$

Try the following exercises.

E5) Show that

$$r_{12\cdot3} = \frac{r_{12} - r_{13}r_{23}}{\left[\left(1-r_{23}^2\right)\left(1-r_{13}^2\right)\right]^{1/2}}.$$

E6) If a random vector $X = (X_1, X_2, X_3, X_4)'$ has covariance matrix

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma^2 & \sigma_{14} & \sigma_{13} \\ \sigma_{13} & \sigma_{14} & \sigma^2 & \sigma_{12} \\ \sigma_{14} & \sigma_{13} & \sigma_{12} & \sigma^2 \end{bmatrix}$$

Then show that the four multiple correlation coefficients between one variable and the other three are equal.

Let us now summarize the unit.

17.6 SUMMARY

In this unit, we have covered the following points:

1. The coefficient of correlation.
2. Multiple correlation is the correlation among more than two variables.
3. Partial correlation is the simple correlation between two variables while the rest of the variables are kept constant.
4. Distribution of $\hat{R}^2$ and $\hat{R}^2/1-\hat{R}^2$.

17.7 SOLUTIONS/ANSWERS

E1) $$x'x = \begin{bmatrix} 12 & 52 & 102 \\ 52 & 395 & 536 \\ 395 & 536 & 1004 \end{bmatrix}$$

$$(x'x)^{-1} = \begin{bmatrix} 0.97 & 0.24 & -0.23 \\ 0.24 & 0.16 & -0.11 \\ -0.23 & -0.11 & -0.08 \end{bmatrix}$$

$$x'y = \begin{bmatrix} 90 \\ 482 \\ 872 \end{bmatrix}$$

$$\hat{b} = (x'x)^{-1} x'y = \begin{bmatrix} 5.38 \\ 3.01 \\ -1.29 \end{bmatrix}$$

$$y = 5.38 + 3.01x_1 - 1.29x_2$$

E2) i) $(I-H)' = \left[I - x(x'x)^{-1}x'\right]'$

$$= I - x\left[\left(x'x\right)^{-1}\right]' x' = H$$

Therefore $I - H$ is symmetric

ii) $$(I-H)^2 = \left[I - x\left(x'x\right)^{-1}x'\right]\left[I - x\left(x'x\right)^{-1}x'\right]$$

$$= I - 2x\left(x'x\right)^{-1}x' + x\left(x'x\right)^{-1}x'x\left(x'x\right)^{-1}x'$$

$$= I - H$$

Therefore, $= I - H$ is idempotent.

E4) i) $z = 4y_1 - 6y_2 + y_3$

$$= \begin{bmatrix} 4 & -6 & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$= a'y$ (suppose)

$$\mu_z = a'\mu = \begin{bmatrix} 4 & -6 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} = 17$$

$$\Sigma_z = a'\Sigma a = \begin{bmatrix} 4 & -6 & 1 \end{bmatrix}\begin{bmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{bmatrix}\begin{bmatrix} 4 \\ -6 \\ 1 \end{bmatrix} = 79$$

Therefore, z is distributed as $N(17, 79)$

$$\mathbf{z} = \begin{bmatrix} y_1 - y_2 + y_3 \\ 2y_1 + y_2 - y_3 \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \text{ (suppose)}$$

$$= \begin{bmatrix} 1 & -1 & 1 \\ 2 & 1 & -1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \mathbf{A}\mathbf{y} \text{ (suppose)}$$

$$\mu_z = A\mu = \begin{bmatrix} 1 & -1 & 1 \\ 2 & 1 & -1 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

$$\Sigma_z = A\Sigma A' = \begin{bmatrix} 5 & 4 \\ 4 & 23 \end{bmatrix}$$

Therefore, the distribution of z is $N_2\left[\begin{bmatrix} 6 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 & 4 \\ 4 & 23 \end{bmatrix}\right]$.

ii) $$f\left(y_2 \mid y_1, y_3\right) = N\left[-\frac{5}{2} + \frac{1}{4}y_1 + \frac{1}{3}y_3, \frac{17}{12}\right]$$

$$f\left(y_2 \mid y_1, y_3\right) = N_2\left[\begin{bmatrix} 2 \\ -2 + \frac{1}{3}y_3 \end{bmatrix}, \begin{bmatrix} 4 & 1 \\ 1 & 5/3 \end{bmatrix}\right].$$

iii) $r_{12} = \frac{\sqrt{2}}{4} = 0.3536$

$r_{12\cdot3} = \sqrt{\frac{3}{20}} = 0.3873.$

E5) We know the following relationship between regression coefficients and simple correlation coefficients.
In this case

$$\hat{a}_1 = r_{13}\left(\frac{\sqrt{\sum y^2}}{\sqrt{\sum x_2^2}}\right) \text{ and } \hat{c}_1 = r_{23}\left(\frac{\sqrt{\sum x_1^2}}{\sqrt{\sum x_2^2}}\right) \qquad \text{(i)}$$

From the correlation coefficients of the two regressions we obtain

$$r_{12}^2 = 1 - \frac{\sum e_1^2}{\sum y^2} = 1 - \frac{\sum y^{*2}}{\sum y^2}$$

and

$$r_{23}^2 = 1 - \frac{\sum e_2^2}{\sum x_1^2} = 1 - \frac{\sum x_1^{*2}}{\sum x_1^2}$$

Therefore

$$\sum y^{*2} = \sum y^2\left(1 - r_{12}^2\right)$$

and

$$\sum x_1^{*2} = \sum x_1^2\left(1 - r_{23}^2\right) \qquad \text{(ii)}$$

Substitute the terms with asterisks in the formula of the partial correlation coefficient

$$r_{12.3} = \frac{\sum(y - \hat{a}_1x_2)(x_1 - \hat{a}_1x_2)}{\sqrt{\sum y^2\left(1 - r_{12}^2\right)}\sqrt{\sum x_1^2\left(1 - r_{23}^2\right)}}$$

$$= \frac{\sum\left(yx_1 - \hat{a}_1x_1x_2 - \hat{c}_1yx_2 + \hat{a}_1\hat{c}_1x_2^2\right)}{\sqrt{\sum y^2 \sum x_1^2}\sqrt{\left(1 - r_{13}^2\right)\left(1 - r_{23}^2\right)}}$$

$$= \frac{\sum yx_1 - \hat{a}_1\sum x_1x_2 - \hat{c}_1\sum yx_2 + \hat{a}_1\hat{c}_1\sum x_2^2}{\sqrt{\sum y^2}\sqrt{\sum x_1^2}\sqrt{\left(1 - r_{13}^2\right)\left(1 - r_{23}^2\right)}} \qquad \text{(iii)}$$

Substituting the values of $\hat{a}_1$ and $\hat{c}_1$ for their expressions obtained in (i)

$r_{12.3} =$

$$\frac{\sum yx_1 - r_{13}\left(\frac{\sqrt{\sum y^2}}{\sqrt{\sum x_2^2}}\right)\sum x_1x_2 - r_{23}\left(\frac{\sqrt{\sum x_1^2}}{\sqrt{\sum x_2^2}}\right)\sum yx_2 + r_{13}r_{23}\left(\frac{\sqrt{\sum y^2 \sum x_1^2}}{\sqrt{\sum x_2^2}}\right)\sum x_2^2}{\sqrt{\sum y^2}\sqrt{\sum x_1^2}\sqrt{\left(1 - r_{13}^2\right)\left(1 - r_{23}^2\right)}}$$

We multiply each term of the numerator by appropriate unitary terms so as to transform them into simple correlation coefficients. For example, we multiply the first terms by $\left[\sqrt{\sum y^2}\sqrt{\sum x_1^2}\Big/\sqrt{\sum y^2}\sqrt{\sum x_1^2}\right]=1$, the second term by $\left[\sqrt{\sum x_1^2}\Big/\sqrt{\sum x_1^2}\right]=1$ and so on.

Thus we obtain

$$r_{12.3} = \frac{\sum yx_1\left(\frac{\sqrt{\sum y^2\sum x_1^2}}{\sqrt{\sum y^2\sum x_1^2}}\right) - r_{13}\sum x_1x_2\left(\frac{\sqrt{\sum y^2\sum x_1^2}}{\sqrt{\sum x_1^2\sum x_2^2}}\right) - r_{23}\left(\frac{\sum yx_2\sqrt{\sum x_1^2\sum y^2}}{\sqrt{\sum y^2}\sqrt{\sum x_2^2}}\right) + r_{13}r_{23}\sqrt{\sum y^2\sum x_1^2}}{\sqrt{\sum y^2}\sqrt{\sum x_1^2}\sqrt{\left(1-r_{13}^2\right)\left(1-r_{23}^2\right)}}$$

$$= \frac{\sqrt{\sum y^2}\sqrt{\sum x_1^2}\left(r_{12} - r_{13}r_{23} - r_{13}r_{23} + r_{13}r_{23}\right)}{\sqrt{\sum y^2}\sqrt{\sum x_1^2}\sqrt{\left(1-r_{13}^2\right)\left(1-r_{23}^2\right)}}$$

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

17.8 PRACTICAL ASSIGNMENT

Session 2

a) Write a program in C-language to fit the model $Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + e_i$ where $i = 1, 2, 3, 4, 5$ using the least square estimates.

b) Write a program in C-language to find the multiple correlation coefficient and the mean square error, if Σ_{xx}, σ_{yy} and σ_{xy} are given.