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

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

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

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

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

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

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

These equations can be written as

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

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

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

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

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

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 $y = xb + e$ (8)

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

i) $E(e_i) = 0$

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

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

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

$$e_i = y_1 - b_0 + b_1 x_{i1} - b_2 x_{i2} \cdots b_k x_{ik}$$

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

$$\mathbf{e'e} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - b_0 x_{i1} - b_1 x_{i2} \cdots b_k x_{ik}\right)^2$$

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

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

$$\hat{e}_i = \hat{y}_i - \hat{b}_0 - \hat{b}_1 x_{i1} + \hat{b}_k x_{ik};\ i = 1, 2, \ldots, n \qquad (10)$$

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

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

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

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

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

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

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

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

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,

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

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

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

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

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

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 $\mathbf{H} = \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'$, then verify that

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

ii) $\hat{e} = \mathbf{I} - \mathbf{H}$ 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 $\mathbf{y}'\mathbf{y}$.

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

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

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

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

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, $b_0 + b_1 x_2 + \cdots + b_k x_k$ to estimate or predict y.

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

Then y is predicted by $b_0 + \mathbf{b}'\mathbf{x}$. Then $e = b_0 - \mathbf{b}'\mathbf{x}$ 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 $b'x = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k$. The multiple correlation coefficient $R_{y.x_1x_2\cdots x_k}$ is thus defined by

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

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

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

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

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 $u'v \le \sqrt{u'u}\sqrt{\vartheta' v}$]

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

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

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

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

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

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

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 $= E\left(y - b_0 - b'x\right)^2$ is minimal. Then it can be proved that minimum mean squared error equals

$\operatorname{var}(y/x) = \sigma_{yy} - \sigma'_{yx}\sum_{xx}^{-1}\sigma_{yx}$. [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 $\operatorname{var}(y/x) = \sigma_y^2\left(1 - \rho^2\right)$,

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

$$= \rho^2$$

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 $x_1, x_2, \ldots, x_k$ is $N_{k+1}(\mu, \Sigma)$ where

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

Now suppose we have sample of size n on ($y, x_1, \ldots, x_k$) 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 $$b_0 + \hat{\mathbf{b}}'\mathbf{x} = \bar{y} + s'_{xy}S_{xx}^{-1}(\mathbf{x} - \bar{x}) \qquad (23)$$

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 Σ_{11}, Σ_{12}, Σ_{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 $z = (Y_1, Y_2, X_1, X_2)'$ 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}, $r_{12\cdot 34}$ and $r_{13\cdot 24}$.

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

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

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

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

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

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

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

$$r_{12\cdot34} = \frac{\sigma_{12\cdot34}}{\sqrt{\sigma_{12\cdot34}\sigma_{22\cdot34}}} = \frac{-24/33}{\sqrt{\frac{126}{33} \cdot \frac{14}{33}}} = \frac{-24}{\sqrt{126 \cdot 14}}$$

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

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 $r_{12\cdot3}$.

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

$$r_{12\cdot3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}}$$

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

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

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

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

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

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

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

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

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

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

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

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

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.

UNIT 18 ORTHOGONAL TRANSFORMATIONS

Structure **Page No**

18.1 INTRODUCTION

You must have studied the linear transformation in your earlier studies. Now in this unit, we shall focus on a particular linear transformation called orthogonal transformation. In Section 18.2, we shall define the orthogonal transformation through a matrix. Then in Section 18.3, we shall construct the orthogonal matrices by using the orthogonality in a matrix. At the end, in the Section 18.4, we shall apply the orthogonal transformation in multivariate normal distribution.

Objectives

After reading this unit, you should be able to:

- define orthogonal transformation through orthogonal matrix;
- learn the properties of orthogonal matrix;
- use the orthogonal transformation for multivariate normal distribution.

18.2 ORTHOGONAL TRANSFORMATION

Let us recall first the linear transformation. For this consider a linear transformation $T : \Omega_X \rightarrow \Omega_Y$ which can be expressed through a matrix A.

That is if we transform X to Y by the transformation T, we can write $Y = T(X) = AX$. Consider that $\Omega_X = \Omega_Y = R^n$ and $Y = AX$ is a linear transformation. If the matrix A happens to be the orthogonal matrix, the transformation is called an orthogonal transformation. It may be noted here that orthogonal matrix is always a square matrix.

Before we proceed let us start with the orthogonal matrices first.

An orthogonal matrix is the real specialization of a unitary matrix, and thus always a normal matrix. Although we consider only real matrices here, the definition can be used for matrices with entries from any field. However, orthogonal matrices arise naturally from inner products, and for matrices of complex numbers that leads instead to the unitary requirement. Orthogonal matrices preserve inner product.

To see the inner product connection, consider a vector X in an *n*-dimensional real inner product space. Written with respect to an orthonormal basis, the squared length of X is $X'X$. If a linear transformation, in matrix form Y = M X, preserves vector lengths, then

$$Y'Y = (MX')(MX) = X'(M'M)X.$$

And it should be equal to $X'X$. This can only happen if $M'M = I$ where I is an identity matrix. Thus, we see that $M' = M^{-1}$ for orthogonal matrix M. You may also note that the columns of an orthogonal matrix are an orthonormal set of vectors. Similarly, the rows of an orthogonal matrix are an orthonormal set of rows. As noted above, in order for a matrix to be orthogonal matrix it must be a square matrix. So a matrix that is not square, but does have orthonormal columns will not be orthogonal. To make the point clear let us consider the following example:

Example 1: Consider the following matrices A and B:

$$A = \begin{bmatrix} 2 & 1 & 3 \\ -1 & 0 & 7 \\ 0 & -1 & -1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & 1 & 1 \\ 1 & -5 & 2 \\ -3 & 1 & -4 \\ 1 & -1 & 1 \end{bmatrix}$$

We know that every matrix with linearly independent columns can be factored as QR where Q is a matrix with orthonormal columns and *R* is an invertible upper triangular matrix.

Thus, A and B can be decomposed as

$$A = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} & \frac{2}{\sqrt{6}} \\ 0 & -\frac{5}{\sqrt{30}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \sqrt{5} & \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ 0 & \frac{6}{\sqrt{30}} & \frac{22}{\sqrt{30}} \\ 0 & 0 & \frac{16}{\sqrt{6}} \end{bmatrix} = Q_1\ R_1 \text{ (say)}$$

$$B = \begin{bmatrix} \frac{2}{\sqrt{15}} & \frac{7}{\sqrt{4485}} & -\frac{118}{\sqrt{20930}} \\ \frac{1}{\sqrt{15}} & -\frac{64}{\sqrt{4485}} & -\frac{11}{\sqrt{20930}} \\ -\frac{3}{\sqrt{15}} & -\frac{18}{\sqrt{4485}} & -\frac{81}{\sqrt{20930}} \\ \frac{1}{\sqrt{15}} & -\frac{4}{\sqrt{4485}} & -\frac{18}{20930} \end{bmatrix} \begin{bmatrix} \sqrt{15} & -\frac{11}{\sqrt{15}} & \frac{17}{\sqrt{15}} \\ 0 & \frac{299}{\sqrt{4485}} & -\frac{53}{\sqrt{4485}} \\ 0 & 0 & \frac{210}{\sqrt{20930}} \end{bmatrix} = Q_2R_2 \text{ (say)}$$

For the matrix A, the matrix Q, given as

$$Q_1 = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} & \frac{2}{\sqrt{6}} \\ 0 & -\frac{5}{\sqrt{30}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

by construction has orthonormal columns and since it is a square matrix it is an orthogonal matrix. But for the matrix B, the matrix

$$Q_2 = \begin{bmatrix} \frac{2}{\sqrt{15}} & \frac{7}{\sqrt{4485}} & -\frac{118}{\sqrt{20930}} \\ \frac{1}{\sqrt{15}} & -\frac{64}{\sqrt{4485}} & -\frac{11}{\sqrt{20930}} \\ -\frac{3}{\sqrt{15}} & -\frac{18}{\sqrt{4485}} & -\frac{81}{\sqrt{20930}} \\ \frac{1}{\sqrt{15}} & -\frac{4}{\sqrt{4485}} & -\frac{18}{\sqrt{20930}} \end{bmatrix}$$

Again by construction has orthonormal columns but it is not a square matrix therefore it is not an orthogonal matrix.

Now, try an exercise.

E1) Find value(s) for a, b, and c for which the following matrix A will be

orthogonal $A = \begin{bmatrix} 0 & -\frac{2}{3} & a \\ \frac{1}{\sqrt{5}} & \frac{2}{3} & b \\ -\frac{2}{\sqrt{5}} & \frac{1}{3} & c \end{bmatrix}$.

You may also deduce from the above discussions that finite-dimensional linear isometries-rotations, reflections, and their combinations-produce orthogonal matrices. The converse is also true: orthogonal matrices imply orthogonal transformations. However, linear algebra includes orthogonal transformations between spaces which may be neither finite-dimensional nor of the same dimension, and these have no orthogonal matrix equivalent.

Orthogonal matrices are important for a number of reasons, both theoretical and practical. The $n \times n$ orthogonal matrices form a group, the orthogonal group denoted by O(n), which, with its subgroups, is widely used in mathematics and the physical sciences. For example, the point group of a molecule is a subgroup of O(3). Because floating point versions of orthogonal matrices have advantageous properties, they are key to many algorithms in numerical linear algebra, such as QR decomposition. As another example, with appropriate normalization the discrete cosine transform (used in MP3 compression) is represented by an orthogonal matrix. Below are a few examples of small orthogonal matrices and possible interpretations:

Example 2:

S. No.	Matrix	Interpretations
1.	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	Identity transformation
2.	$\begin{bmatrix} 0.96 & -0.28 \\ 0.28 & 0.96 \end{bmatrix}$	Rotation by 16.26^0
3.	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	Reflection across x-axis

Now, try an exercise.

E2) Check whether the following matrix A has roto inversing or reflection across x-axis or y-axis.

$$A = \begin{bmatrix} 0 & -0.80 & -0.60 \\ -0.80 & -0.36 & 0.48 \\ 0.60 & 0.48 & -0.64 \end{bmatrix}.$$

Now, in the following section we shall discuss the methods of construction of orthogonal transformation.

18.3 CONSTRUCTION OF ORTHOGONAL TRANSFORMATION

In this section, we will discuss briefly the method of construction of orthogonal transformations. We have seen in the previous section that an orthogonal transformation Y = AX can be constructed through an orthogonal matrix A. Now, we will see how we can develop an orthogonal matrix of A of given dimension.

Let us start with the matrix of order 1 that is one dimensional.

The simplest orthogonal matrices are the 1×1 matrices namely $[1]$ and $[-1]$ which we can be interpreted as the identity and a reflection of the real line across the origin.

Consider the case of two dimensional arrays that is the matrix of order 2×2. The 2×2 matrices have the form

$$\begin{bmatrix} p & t \\ q & u \end{bmatrix}.$$

The orthogonality demands to satisfy the following three equations:

$$1 = p^2 + q^2$$
$$1 = t^2 + u^2$$
$$0 = pt + qu.$$

Since there are four unknowns and three equations, more than one solution exists. Let us consider that $p = \cos\theta$, $q = \sin\theta$; then either $t = -q$, $u = p$ or $t = q$, $u = -p$. We can interpret the first case as a rotation by θ (where $\theta = 0$ is the identity), and the second as a reflection across a line at an angle of $\theta/2$. Thus, we see that

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

is an orthogonal transformation which is a rotational transformation. Similarly,

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

is an orthogonal transformation which is a reflectional transformation.

The reflection at 45° exchanges X_1 and X_2; it is a permutation matrix, with a single 1 in each column and row and 0 otherwise, i.e.

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

The identity is also a permutation matrix.

A reflection is its own inverse, which implies that a reflection matrix is symmetric (equal to its transpose) as well as orthogonal. The product of two rotation matrices is a rotation matrix, and the product of two reflection matrices is also a rotation matrix.

Now, let us discuss this for matrices of higher dimension.

Regardless of the dimension, it is always possible to classify orthogonal matrices as purely rotational or not, but for 3×3 matrices and larger the non-rotational matrices can be more complicated than reflections. For example,

$$\begin{bmatrix} -1 & 02 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

represents an *inversion* through the origin and a *rotoinversion* about the z-axis. Rotations also become more complicated; they can no longer be completely characterized by an angle, and may affect more than one planar subspace. While it is common to describe a 3×3 rotation matrix in terms of an axis and angle, the existence of an axis is an accidental property of this dimension that applies in no other. However, we have elementary building blocks for permutations, reflections, and rotations that apply in general.

Primitives

The most elementary permutation is a transposition, obtained from the identity matrix by exchanging two rows. Any $n\times n$ permutation matrix can be constructed as a product of at most $n-1$ transpositions.

A Householder reflection is constructed from a non-null vector $\mathbf{v}$. Here the numerator is a symmetric matrix while the denominator is a number, the squared magnitude of v. This is a reflection in the hyperplane perpendicular to v (negating any vector component parallel to v). If v is a unit vector, then $Q = I - 2\mathbf{v}\mathbf{v}^T$ suffices. A Householder reflection is typically used to simultaneously zero the lower part of a column. Any orthogonal matrix of size $n\times n$ can be constructed as a product of at most n such reflections.

A given rotation acts on a two-dimensional (planar) subspace spanned by two coordinate axes, rotating by a chosen angle. It is typically used to zero a single sub diagonal entry. Any rotation matrix of size $n\times n$ can be constructed as a product of at most $n(n-1)/2$ such rotations. In the case of 3×3 matrices, three such rotations suffice; and by fixing the sequence we can thus describe all 3×3 rotation matrices (though not uniquely) in terms of the three angles used, often called Euler angles. A Jacobi rotation has the same form as a Givens rotation, but is used as a similarity transformation chosen to zero both off-diagonal entries of a 2×2 symmetric submatrix.

Matrix Properties of Orthogonality

A real square matrix is orthogonal if and only if its columns form an orthonormal basis of the Euclidean space $\mathbf{R}^n$ with the ordinary Euclidean dot product, which is the case if and only if its rows form an orthonormal basis of $\mathbf{R}^n$. It might be tempting to suppose a matrix with orthogonal (not orthonormal) columns would be called an orthogonal matrix, but such matrices have no special interest and no special name; they only satisfy $M^TM = D$, with D a diagonal matrix.

The determinant of any orthogonal matrix is $+1$ or -1. This follows from basic facts about determinants, as follows:

$$1 = \det(I) = \det(Q^tQ) = \det(Q^t)\det(Q) = (\det(Q))^2.$$

The converse is not true; having a determinant of $+1$ is no guarantee of orthogonality, even with orthogonal columns, as shown by the following counter example:

$$\begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}.$$

The inverse of every orthogonal matrix is again orthogonal, as is the matrix product of two orthogonal matrices.

Example 3: Construct an orthogonal matrix of order 3×3 such that the elements in the first row are equal.

Solution: Let us suppose that the elements in the first row are 'a'. In order that the matrix is orthogonal, sum of square of these must be one, i.e. $3a^2 = 1$. This shows that the matrix may be of the following forms:

$$\begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ b & d & f \\ c & e & g \end{bmatrix} \text{ or } \begin{bmatrix} -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \\ b & d & f \\ c & e & g \end{bmatrix} \text{ or } \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ b & d & f \\ c & e & g \end{bmatrix} \text{ and so on.}$$

Let us consider the first one. In order that it is an orthogonal matrix rest of the elements should fulfill the following conditions:

$$b + d + f = 0, \tag{1}$$

$$c + e + g = 0, \tag{2}$$

$$bc + de + fg = 0, \tag{3}$$

$$b^2 + d^2 + f^2 = 1, \tag{4}$$

$$c^2 + e^2 + g^2 = 1. \tag{5}$$

If we choose $b = 2/\sqrt{6}$, $d = -1/\sqrt{6}$ and $f = -1/\sqrt{6}$, Eqn. (1) and Eqn. (4) are satisfied. Eqn. (3) reduces to the following on substitution of these values:

$$2c - e - g = 0 \tag{6}$$

On adding it with Eqn. (2) we get $c = 0$ and hence $e = -g$. From Eqn. (5) we have $e = -g = \pm 1/\sqrt{2}$. Therefore all of the following will be an orthogonal matrix.

$$\begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \text{ or } \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Similarly, you consider other forms and develop orthogonal matrices. Hence show that many more orthogonal matrices can be constructed.

So far, we have discussed the orthogonal matrices in detail. Now in the following section, let us focus on the orthogonal transformation in multivariate normal distribution.

18.4 USE OF ORTHOGONAL TRANSFORMATION IN MULTIVARIATE NORMAL DISTRIBUTIONS

Let X be a p-dimensional random vector such that $X = [X_1, X_2, \ldots, X_n]^t$ and the density function of X, i.e. the joint density function of $X_1, X_2, \ldots, X_p$ be $f(x_1, x_2, \ldots, x_p)$. Let us define a real valued function $Y = y(x)$ i.e.

$$Y_j = y_j(x_1, x_2, \ldots, x_p); \; j = 1, 2, \ldots, p. \tag{7}$$

If the transformation X to Y is one-to-one, the inverse transformation can be written as

$$X_i = x_i(y_1, y_2, \ldots, y_p). \tag{8}$$

Now, you know that the density function of random vector Y, i.e. the joint density of $Y_1, Y_2, \ldots, Y_p$ is obtained as

$$\begin{aligned} G(y_1, y_2, \ldots, y_p) &= f(x_1(y_1, y_2, \ldots, y_p), x_2(y_1, y_2, \ldots, y_p) \\ &\quad \ldots x_i(y_1, y_2, \ldots, y_p)).|J(y_1, y_2, \ldots, y_p)|. \end{aligned} \tag{9}$$

where $|J|$ is the magnitude (positive value) of the Jacobian of the transformation defined below:

$$J(y_1, y_2, \ldots, y_p) = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \cdots & \frac{\partial x_1}{\partial y_p} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \cdots & \frac{\partial x_2}{\partial y_p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial x_p}{\partial y_1} & \frac{\partial x_1}{\partial y_1} & \cdots & \cdots & \frac{\partial x_p}{\partial y_p} \end{vmatrix}. \tag{10}$$

Theorem 1: The Jacobian of an orthogonal transformation is always one.

Proof: Let us consider that a p-dimensional random vector X is transformed to p-dimensional random vector Y by an orthogonal transformation $Y = AX$, where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix}. \tag{11}$$

Thus in the expanded form the transformation is written as

$$\begin{aligned} Y_1 &= a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\ Y_2 &= a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ Y_p &= a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p \end{aligned} \tag{12}$$

Now the inverse transformation can easily be obtained by recalling that $Y = AX$ is an orthogonal transformation and A is an orthogonal matrix. Thus, $A^tA = I$. Pre-multiplying by A^t in $Y = AX$ on both the sides, we have $A^tY = A^tAX = X$, therefore the inverse transformation in extended form can be written as

$$\begin{aligned} X_1 &= a_{11}Y_1 + a_{21}Y_2 + \cdots + a_{p1}Y_p \\ X_2 &= a_{12}Y_1 + a_{22}Y_2 + \cdots + a_{p2}Y_p \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ X_p &= a_{1p}Y_1 + a_{2p}Y_2 + \cdots + a_{pp}Y_p \end{aligned} \tag{13}$$

Now, using Eqn. (10), we get the Jacobian of the transformation as

$$J = \begin{vmatrix} a_{11} & a_{21} & \dots & \dots & a_{p1} \\ a_{12} & a_{22} & \dots & \dots & a_{p2} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ a_{1p} & a_{2p} & \dots & \dots & a_{pp} \end{vmatrix} = |A^t| \tag{14}$$

But $A^tA = I$, therefore $|A^tA| = 1 \Rightarrow |A^t||A| = 1$. The magnitude of the determinant of A and A^t are same and hence magnitude of $|A^t|$ = magnitude of $|A| = 1$.

Remark: From the above theorem we see that if we have to use orthogonal transformation, there is in fact no need to calculate the Jacobian of the transformation. The density of the transformed new random variable can simply be obtained by replacing the old variables in terms of new variable as obtained from the inverse transformation. This ease of the orthogonal transformation gives the following interesting result which is popularly known as **Fisher's lemma.**

Theorem 2: Let X is p component random vector following normal distribution with mean vector 0 and variance covariance matrix I (now onwards it will be denoted as $X \sim N_p(0, I)$. Then every orthogonal transformation $Y = AX$ produces a new random vector Y having similar distribution, i.e. $Y \sim N_p(0, I)$.

Proof: Since $X \sim N_p(0, I)$, the density function of x is given as:

$$f(X) = (2\pi)^{-p/2} \exp(-X^tX) \tag{15}$$

Hence the density of $Y = AX$, where A is orthogonal matrix can be easily obtained, using the above theorem by replacing X by A^tY in (16), as

$$\begin{aligned} g(Y) &= (2\pi)^{-p/2} \exp(-(A^tY)^t(A^tY)) \\ &= (2\pi)^{-p/2} \exp(-Y^tA^tAY) \\ &= (2\pi)^{-p/2} \exp(-Y^tY) \qquad \text{(since } A^tA = I) \end{aligned} \tag{16}$$

Remark: The Eqn. (15) can be written as

$$f(X_1, X_2, \dots, X_p) = (2\pi)^{-p/2} \exp\left[-\left(\sum_{i=1}^{p} X_i^2\right) = \prod_{i=1}^{p} \frac{1}{2\pi} \exp(X_i^2)\right] \tag{17}$$

This means that X_i's are independently and identically distributed (I.I.D.) standard normal variables. Thus, the theorem can also be stated as "Orthogonal transformation on a set of I.I.D. standard normal variables gives a new set of I.I.D. standard normal variables".

Let us consider another application of orthogonal transformation in studying the effect of independent linear constraints on chi-square distribution. Let X be an n-dimensional random vector such that $X \sim N_n(0, I)$, i.e. components X_i's are I.I.D. standard normal variables. Therefore you know that $X^tX = \sum_{i=1}^{n} X_i^2$ is distributed as chi-square with n degrees of freedom. However, if the variables are subject to linear constraints, the degrees of freedom of X^tX is reduced.

Theorem 3: If $X \sim N_n(0, I)$ where X is subjected to m linearly independent constraints $BX = 0$, the distribution of X^tX is chi-square with $n - m$ degrees of freedom.

Proof: Let us consider the expanded form of the constraints given below:

$$\begin{aligned} &b_{11}X_1 + b_{12}X_2 + \cdots + b_{1n}X_n = 0 \\ &b_{21}X_1 + b_{22}X_2 + \cdots + b_{2n}X_n = 0 \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &b_{m1}X_1 + b_{m2}X_2 + \cdots + b_{mn}X_n = 0 \end{aligned} \qquad (18)$$

Consider a transformation $Y = AX$ such that jth element in the first row of the matrix A is

$$a_{1j} = \frac{b_{1j}}{\sqrt{\sum_{i=1}^{n} b_{1j}^2}} \text{ for } j = 1, 2, \ldots, n. \qquad (19)$$

Now, we choose rest $n(n-1)$ of the elements in A such that it is an orthogonal matrix, i.e. we have $n(n-1))$ unknowns to be obtained by solving $(n-1)(n-2)/2$ equations for $n(n-1)$ unknowns. Since the number of unknowns is more than the number of equations; more than one solution exists. Further,

$$Y_1 = \sum_{j=1}^{n} a_{1j}X_j = \frac{\sum_{j=1}^{n} b_{1j}X_j}{\sqrt{\sum_{j=1}^{n} b_{1j}^2}} = 0, \qquad (20)$$

because of the first linear constraints. It may also be noted that
$Y^tY = (AX)^t(AX) = X^tA^tAX = X^tX$, i.e.

$$\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} X_i^2 = \sum_{i=2}^{n} Y_i^2 \qquad (Y_1 = 0). \qquad (21)$$

In terms of Y_i's the constraints (18) reduces to $0 = BX = BA^tY = CY$, i.e.

$$\begin{bmatrix} b_{11} & b_{12} & \cdots & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{m1} & b_{m2} & \cdots & \cdots & b_{mn} \end{bmatrix} \begin{bmatrix} \frac{b_{11}}{\sqrt{\sum_{i=1}^{n} b_{1j}^2}} & \frac{b_{12}}{\sqrt{\sum_{i=1}^{n} b_{1j}^2}} & \cdots & \cdots & \frac{b_{1n}}{\sqrt{\sum_{i=1}^{n} b_{1j}^2}} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nn} \end{bmatrix}^t \begin{bmatrix} 0 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

or,

$$\begin{bmatrix} b_{11} & b_{12} & \cdots & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{m1} & b_{m2} & \cdots & \cdots & b_{mn} \end{bmatrix} \begin{bmatrix} \frac{b_{11}}{\sqrt{\sum_{i=1}^{n} b_{1j}^2}} & a_{21} & \cdots & \cdots & a_{n1} \\ \frac{b_{12}}{\sqrt{\sum_{i=1}^{n} b_{1j}^2}} & a_{22} & \cdots & \cdots & a_{n2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{b_{1n}}{\sqrt{\sum_{i=1}^{n} b_{1j}^2}} & a_{2n} & \cdots & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} 0 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

or,

$$\begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ c_{21} & c_{22} & \cdots & \cdots & c_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{m1} & c_{m2} & \cdots & \cdots & c_{mn} \end{bmatrix} \begin{bmatrix} 0 \\ Y_2 \\ \vdots \\ \vdots \\ Y_n \end{bmatrix} = 0$$

or,

$$\begin{aligned} c_{22}Y_2 + c_{23}Y_3 + \cdots + c_{2n}Y_n &= 0 \\ c_{32}Y_2 + c_{33}Y_3 + \cdots + c_{3n}Y_n &= 0 \\ &\cdots\cdots \\ &\cdots\cdots \\ c_{m2}Y_2 + c_{m3}Y_3 + \cdots + c_{mn}Y_n &= 0 \end{aligned} \qquad (22)$$

Therefore, the original problem of finding the distribution of $\sum_{i=1}^{n} X_i^2$ (sum of n standard normal variables) subject to the constraint $BX = 0$, now reduces to finding the distribution of $\sum_{i=2}^{n} Y_i^2$ (sum of $(n-1)$ standard normal variables) subject to $(m-1)$ linear constraints. Needless to mention that the transformation X to Y be orthogonal therefore Y_i's are also standard normal variables (as per Theorem (2)).

Now we repeat the above process through transforming Y to Z using an orthogonal transformation $Z = A^* Y$ (where $A^* = ((a_{ij}))$; i, J = 2, 3, ...n) and choosing

$$a_{2j} = \frac{c_{2j}}{\sqrt{\sum_{j=2}^{n} c_{2j}^2}}.$$

We see that original problem of finding the distribution of sum of n standard normal variables subject to m linear constraints, now reduces to finding the distribution of sum of $(n-2)$ standard normal variables subject to $(m-2)$ linear constraints. Repeating this process for m times we conclude that the problem finally reduces to find the distribution of sum of square of $(n-m)$ standard normal variables free from any constraint thus follows chi-square with $(n-m)$ degrees of freedom.

Now, we shall summarize the unit.

18.5 SUMMARY

In this unit, we have covered the following points.

1. We introduced the concepts transformation particularly orthogonal transformation. Also the advantages of orthogonal transformation were discussed.
2. We generated orthogonal matrices of our interest for manipulating the problems in finding the distributions. The examples given in the unit were deliberately taken to be of low dimension; however they will learn the procedure and may use it for higher dimension also.
3. The orthogonal transformation of multivariate normal distribution was discussed in detail.

18.6 SOLUTIONS/ANSWERS

E1) You see that the columns of the matrix M are

$$M_1 = \begin{bmatrix} 0 \\ \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{bmatrix} \quad M_2 = \begin{bmatrix} -\frac{2}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{bmatrix} \quad M_3 = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

You can verify that M_1 and M_2 are orthonormal vectors possess the properties that $M_1.M_1 = 1$, $M_2.M_2 = 1$ and $M_1.M_2 = 0$, i.e. these fulfill the requirement of columns of an orthogonal matrix. Therefore, we need to find a, b and c such that $M_3.M_3 = 1$, $M_1.M_3 = 0$ and $M_2.M_3 = 0$, i.e.

$$M_1.M_3 = \frac{b}{\sqrt{5}} - \frac{2c}{\sqrt{5}} = 0$$

$$M_2.M_3 = -\frac{2}{3}a + \frac{2}{3}b + \frac{1}{3}c = 0$$

From the first dot product we can see that $b = 2c$. Substituting this into the second dot product, we get $5c = 2a$. Combining these results we have the values of b and c in terms of a as $b = 4a/5$ and $c = 2a/5$. Now, in order that the third column be orthonormal to make the matrix M orthogonal, we must have $M_3.M_3 = 1$.

Since $M_3 = \begin{bmatrix} a \\ \frac{4}{5}a \\ \frac{2}{5}a \end{bmatrix}$ $M_3.M_3 = a^2 + \frac{16}{25}a^2 + \frac{4}{25}a^2 = 1$

This gives us two possible values of a namely $a = \pm\frac{5}{\sqrt{45}}$ that we can use and this in turn means that we could used either of the following two vectors

$$M_3 = \begin{bmatrix} \frac{5}{\sqrt{45}} \\ \frac{4}{\sqrt{45}} \\ \frac{2}{\sqrt{45}} \end{bmatrix} \text{ or, } M_3 = \begin{bmatrix} -\frac{5}{\sqrt{45}} \\ -\frac{4}{\sqrt{45}} \\ -\frac{2}{\sqrt{45}} \end{bmatrix}$$

E2)

S.No.	Matrix	Interpretations
1.	$\begin{bmatrix} 0 & -0.80 & -0.60 \\ -0.80 & -0.36 & 0.48 \\ 0.60 & 0.48 & -0.64 \end{bmatrix}$	Rotoinversion: axis (0, -3/5, 4/5), angle 90^0

18.7 PRACTICAL ASSIGNMENT

1) Write a program in 'C' language to show that the matrix A is orthogonal.

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

Also, extend the program to check that the columns of A form an orthonormal basis of R^4.

2) **Write a program in 'C' language to develop an orthogonal matrix of order 3×3 such that the elements in the first row are equal.**

UNIT 19 INFERENCE-I

Structure **Page No**

19.1 INTRODUCTION

We shall begin this unit by recalling the multivariate normal distribution. Many of our explanations use the representation of the rows of X as k points in n-dimensions. In section 19.2, we will introduce the assumptions that a random sample is constituted by the observations. We shall confine it for random sampling, for which we shall assume that the traits or the measurements considered for the different trials are independent and the joint distribution of all the variable remains the same. In section 19.3, we shall discuss the maximum likelihood estimation.

Objectives

After studying this unit, you should be able to:

- describe the likelihood function and the role of maximum likelihood estimation in deriving the normally distributed estimators;
- understand the distribution of a normal random vector;
- compute the mean and covariance;
- assess the role of multivariate normal distribution in finance.

19.2 MULTIVARIATE NORMAL DISTRIBUTION AND RANDOM SAMPLING

Let us recall the multivariate normal distribution, A p-dimensional random variable x is said to have a p variate normal distribution if and only if every linear function of x has a univariate normal distribution.

Let $X' = (X_1, X_2, \ldots, X_p)$ represent a p-dimensional random variable. The mean and variance-covariance matrix are denoted by μ and Σ, respectively and are given by

$$\mu' = E(x') = (E(x_1), E(x_2), \ldots, E(x_p)) = (\mu_1, \mu_2, \ldots, \mu_p) \qquad (1)$$

$$\Sigma = E[(x-\mu)\,(x-\mu)] \qquad (2)$$

$$= \begin{bmatrix} var(x_1) & cov(x_1, x_2) & \ldots & cov(x_1, x_2) \\ cov(x_2, x_1) & var(x_2) & \ldots & cov(x_2, x_p) \\ \vdots & \vdots & \ldots & \vdots \\ cov(x_p, x_1) & cov(x_p, x_2) & \ldots & var(x_p) \end{bmatrix}$$

The density of a p-variate normal random variable x is given by

$$N_p(\mu, \Sigma) = (2\pi)^{-1/2}\,|\Sigma|^{-1/2}\exp\left\{\frac{-1}{2}(x-\mu)'\,\Sigma^{-1}(x-\mu)\right\} \qquad (3)$$

where the mean of x is μ, and the variance-covariance matrix of x is Σ.

We now obtain the multivariate normal density by transforming the random vector $z = (z_1, z_2, \ldots, z_p)$ where each z_i has the $N(0,1)$ and the z_is are independent. Thus $E(z) = 0$ and $\text{cov}(z) = I$. When random variables are independently distributed, then their density function is the product of their individual densities. We can write

$$f(z) = f_1(z_1)\, f_2(z_2) \ldots f_p(z_p)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-z_1^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-z_2^2/2} \cdots \frac{1}{\sqrt{2\pi}} e^{-z_p^2/2} \left[\text{since } f_i(z_i) = \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2}\right]$$

$$= \frac{1}{\left(\sqrt{2\pi}\right)^p} . e^{\sum_j z_j^2/2}$$

$$= \frac{1}{(\sqrt{2\pi})^p} e^{-z'z/2} \tag{4}$$

If we extend z as a multivariate normal density, then it is with mean vector **0** and covariance matrix **I**.

Here, we wish to obtain the multivariate normal density with arbitrary mean vector μ and covariance matrix Σ which is positive definite. We define the transformation for $y = \sigma z + \mu$, $y = \Sigma^{1/2} z + \mu$.

where $\Sigma^{1/2}$ is the (symmetric) square root matrix. The mean vector and covariance matrix for the transformed random vector y are :

$$E(y) = E\left(\Sigma^{1/2} z + \mu\right) = \Sigma^{1/2} E(z) + E(\mu)$$

$$= \Sigma^{1/2} \cdot 0 + \mu = \mu \tag{5}$$

$$\text{cov}(y) = \text{cov}\left(\Sigma^{1/2} z + \mu\right) = \Sigma^{1/2} \text{cov}(z) \left(\Sigma^{1/2}\right)'$$

$$= \Sigma^{1/2} \cdot I \cdot \Sigma^{1/2} = \Sigma \tag{6}$$

Let us now find the density of $y = \Sigma^{1/2} z + \mu$ from the density of z given in Eqn. (4).

The density of $y = \sigma z + \mu$ is $g(y) = f(z) \frac{dz}{dy} = f(z) |1/\sigma|$.

The analogous expression for

$$y = \Sigma^{1/2} z + \mu \text{ is } g(y) = f(z) \left|\Sigma^{1/2}\right| \tag{7}$$

It may be noted that in Eqn. (7) we use the vertical bars for the determinant of a matrix as opposed to the use of vertical bars to indicate the absolute value of a scalar quantity used as $\left|\frac{dz}{dy}\right|$.

Now Eqn. (7) can be written as

$$g(y) = f(z) \left|\Sigma^{-1/2}\right| \tag{8}$$

$$= f(z) |\Sigma|^{-1/2} \tag{9}$$

$$= \frac{1}{\left(\sqrt{2\pi}\right)^p} \cdot \frac{1}{|\Sigma|^{1/2}} e^{-z'z/2} \qquad \text{[using Eqn. (4)]}$$

$$= \frac{1}{\left(\sqrt{2\pi}\right)^{p} |\Sigma|^{1/2}} e^{-\left[\Sigma^{-1/2}(y-\mu)\right]'\left[\Sigma^{-1/2}(y-\mu)\right]/2} \qquad \left[\text{since } Z = \frac{y-\mu}{\Sigma^{1/2}}\right]$$

$$= \frac{1}{\left(\sqrt{2\pi}\right)^{p} |\Sigma|^{1/2}} e^{-(y-\mu)'\left(\Sigma^{1/2}\Sigma^{1/2}\right)^{-1}(y-u)/2}$$

$$= \frac{1}{\left(\sqrt{2\pi}\right)^{p} |\Sigma|^{1/2}} e^{-(y-\mu)'\Sigma^{-1}(y-\mu)/2} \tag{10}$$

Which is the multivariate normal density function with mean vector u and covariance matrix Σ.

Now, we shall discuss random sampling. For this, consider the data matrix X of order $n \times k$, which can be plotted for the n-dimensional scatterplot by representing the columns of X as points. We can write

$$X_{n\times k} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} = \left[x_1, x_2, \cdots x_n\right] = \begin{bmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_n \end{bmatrix}$$

where y'_i is considered as the elements of the rows of the data matrix. Let us consider $\overline{x_i}$ as the sample mean of the i-th observation which is $\overline{x_i} = \frac{1}{n}\sum_{h=1}^{n} x_{hi}$; $i = 1, 2, \ldots, k$

then the deviation vector denoted by e_i is computed as

$$e_i = y_i - \overline{x}_i 1, \text{ where } 1 = [1, 1, \ldots, 1]'$$

$$= [x_{i1} - \overline{x}_i, x_{i2} - \overline{x}_i \ldots x_{ik} - \overline{x}_i]'$$

for two vectors e_i and e_j, we have $e_i'e_j = n\, s_{ij}$, where s_{ij} is the sample variance-covariance matrix. Also, the sample correlation coefficient

$$\hat{R} = \begin{bmatrix} 1 & r_{iy} \\ r_{iy} & 1 \end{bmatrix}, \text{ with } r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}} = \cos\left(\theta_{ij}\right)$$

where θ_{ij} is the angle between e_i and e_j.

Example 1: For $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$

With $\sigma_1^2, \sigma_2^2 > 0, -1 < \rho < 1$.

Find the expression for the probability density function of a bivariate normal distribution.

Solution: Since $\det\Sigma = \sigma_1^2\sigma_2^2\left(1-\rho^2\right) > 0$,

Σ^{-1} exists and is given by

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_1^2} \end{bmatrix} \cdot \frac{1}{1-\rho^2}$$

Furthermore, for $X = \left(x_1, x_2\right)' \neq 0$

$$X'\Sigma X = \left(\sigma_1 x_1 + \rho\sigma_2 x_2\right)^2 + \left(1-\rho^2\right)\sigma_2^2 x_2^2 > 0.$$

Hence Σ is positive definite.

With $\mu = (\mu_1, \mu_2)'$

$$(x-\mu)'\Sigma^{-1}(x-\mu) = \frac{1}{1-\sigma^2}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right).$$

The probability density function of a bivariate normal random variable with values in E^2 is

$$\frac{1}{2\pi\sigma_1\sigma_2\left(1-\rho^2\right)^{1/2}}\exp\left\{-\frac{1}{2\left(1-\rho^2\right)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right\}.$$

Example 2: Let $\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

Since, for $X = (x_1, x_2)' \neq 0$

$$X'\Sigma X = 2(x_1 + x_2/2)^2 + \frac{3}{2}x_2^2 > 0,$$

Σ is positive definite. Hence

$$f_x(x) = \frac{1}{2\pi\sqrt{3}}\exp\left\{-\frac{2}{3}\left[\frac{(x_1-2)^2}{2} + \frac{(x_2-3)^2}{2} - \frac{(x_1-2)(x_2-3)}{2}\right]\right\}$$

is the probability density function of a bivariate normal random variable $X = (X_1, X_2)'$ with mean $(2,3)'$ and covariance Σ.

Here $\rho = \frac{1}{2}$.

Example 3 (The equivalence of zero covariance and independence for normal variables): Let $X_{(3\times1)}$ be $N_3(\mu, \Sigma)$ with

$$\Sigma = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Check whether X_1 and X_2 are independent or not? Also check the independence of (X_1, X_2) and X_3?

Since X_1 & X_2 have covariance $\sigma_{12} = 1$, they are not independent. However, partitioning X & Σ as

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \cdots \\ X_3 \end{pmatrix}, \Sigma = \begin{bmatrix} 4 & 1 & \vdots & 0 \\ 1 & 3 & \vdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \vdots & 2 \end{bmatrix} = \begin{bmatrix} \underset{2\times2}{\Sigma_{11}} & \vdots & \underset{2\times1}{\Sigma_{12}} \\ \cdots & \cdots & \cdots \\ \underset{1\times2}{\Sigma_{21}} & \vdots & \underset{1\times1}{\Sigma_{22}} \end{bmatrix}$$

We see that $X_1 = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ and X_3 have covariance matrix $\Sigma_{12} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Therefore, (X_1, X_2) and X_3 are independent. This implies X_3 is independent of X_1 and also of X_2.

Example 4 (The distribution of a linear combination of the components of a normal random vector): Consider the linear combination $a'x$ of a multivariate normal random vector determined by the choice $a'x = [1,0,\cdots,0]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = x_1$.

And

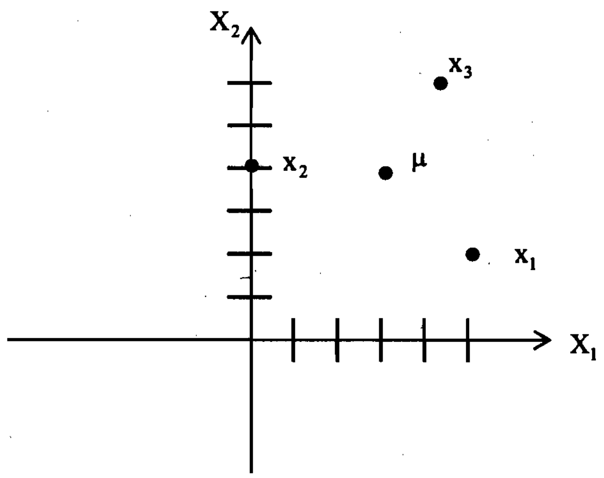
$$a'\mu = [1,0,\cdots,0]\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \mu_1.$$

We have

$$a'\Sigma a = [1,0,\cdots,0]\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \sigma_{11}$$

and it follows from Result (1) that X_1 is distributed as $N(\mu_1, \sigma_{11})$. More generally, the marginal distribution of any component X_i of X is $N(\mu_i, \sigma_{ii})$.
[**Result 1:** If X is distributed as $N_p(\mu, \Sigma)$, then any linear combination of variables $a'X = a_1X_1 + a_2X_2 + \cdots + a_p \times p$ is distributed as $N(a'\mu, a'\Sigma a)$.]

Example 5: Consider $X = \begin{bmatrix} 5 & 0 & 4 \\ 2 & 4 & 6 \end{bmatrix}$, the scatter plot of X is



Mean vector is $\mu = \begin{bmatrix} \dfrac{5+0+4}{3} \\ \dfrac{2+4+6}{3} \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$

Here, $e_1 = y_1 - \mu_1 1 = \begin{bmatrix} 5 \\ 0 \\ 4 \end{bmatrix} - \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}$

$$e_2 = y_2 - \mu_2 1 = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix}$$

Also, $3s_{11} = e_1'e_1 = 14$

$3s_{12} = e_1'e_2 = -2$

$3s_{22} = e_2'e_2 = 8$

Therefore, $s_n = \begin{bmatrix} \frac{14}{3} & -2/3 \\ -2/3 & 8/3 \end{bmatrix}, r_{12} = \frac{s_{12}}{\sqrt{s_{11}s_{22}}} = -0.189$

and $\hat{R} = \begin{bmatrix} 1 & -0.189 \\ -0.189 & 1 \end{bmatrix}$.

Now, try the following exercises:

E1) For X distributed as $N_3(\mu, \Sigma)$, Find the distribution of $\begin{bmatrix} X_1 & - & X_2 \\ X_2 & - & X_3 \end{bmatrix}$.

E2) If X is distributed as $N_5(\mu, \Sigma)$, find the distribution of $\begin{bmatrix} X_2 \\ X_4 \end{bmatrix}$.

E3) Prove that if the covariance matrix Σ of a normal random vector $X = (X_1, \cdots, X_p)'$ is a diagonal matrix, then the components of X are independently distributed normal random variables.

E4) Find the sample correlation matrix $\hat{R}$ for the data matrix $X = \begin{bmatrix} -1 & 2 & 5 \\ 3 & 4 & 2 \\ -2 & 2 & 3 \end{bmatrix}$.

E5) Let X be $N_3(\mu, \Sigma)$ with $\mu' = [-3, 1, 4]$ and $\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$. Which of the following random variables are independent? Give reasons.

a) X_1 and X_2

b) X_2 and X_3

c) (X_1, X_2) and X_3

d) $\left(\frac{X_1 + X_2}{2}\right)$ and X_3

e) X_2 and $\left(X_2 - \frac{5}{2}X_1 - X_3\right)$.

E6) Find the mean and the covariance matrix of the random vector $X = (X_1, X_2)'$ with probability density function

$f_x(X) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left(2x_1^2 + x_2^2 + 2x_1x_2 - 22x_1 - 14x_2 + 65\right)\right\}$ and $X \in E^2$.

In the following section, we shall define and maximum likelihood estimation of multivariate normals.

19.3 MAXIMUM LIKELIHOOD ESTIMATION

Suppose $x_1, x_2, \cdots, x_n$ is a random sample of vector observation from a population with a density $h(x_j;\theta)$, where θ is a parameter vector that identifies the particular member of a family being sampled.
The likelihood function, denoted by $L(x;\theta)$, is defined by

$$L(x;\theta) = \prod_{i=1}^{n} h(x_i;\theta)$$

where we view $L(x;\theta)$ as a function of θ for a fixed set of sample observation $x = (x_1, x_2, \cdots, x_n)$.
In the case where the density is $N_p(\mu, \Sigma)$, then

$$L(x, \mu, \Sigma) = \left(\frac{1}{2\pi}\right)^{n/2} |\Sigma|^{-n/2} \exp\left\{\frac{-1}{2}\sum_i (x_i - \mu)\Sigma^{-1}(x_i - \mu)\right\}.$$

Frequently, the logarithm of the likelihood function is used.

The value of a parameter, as a function of the data, that maximizes the likelihood function is called a maximum likelihood estimator (MLE). In many important cases MLEs can be found quite readily by differentiating the likelihood function, or its logarithm, setting the result equal to zero, and solving. In other cases the maximum occurs at a point where the derivative does not exist, and hence other procedures have to be followed. Often numerical methods have to be used. Iterative proportional fitting and the Newton-Raphson method are two such algorithms frequently used to generate MLEs in this case.

Properties of Maximum Likelihood Estimators

1) **Unbiasedness:** An estimate S of a parameter θ is said to be unbiased if $E(S) = \theta$. In the case where S is a vector or matrix of estimators, then we say the estimate is unbiased if the expected values of the constituent components equal the component values of the vector or matrix of parameters. The sample mean $\overline{X}$ is an unbiased estimate of the population mean vector μ, and if the sample sums-of-squares cross-product matrix is divided by the sample size minus 1, then it is an unbiased estimate of the true variance-covariance matrix.

2) **Sufficiency:** Suppose that the data matrix $X = (X_1, X_2, \ldots, X_n)$ where X_i is a p-dimensional vector, forms a realization of a random vector, and that a family F of possible distributions is specified. A statistic S is said to be sufficient for F if the conditioned density $f_{XIS}(XIS)$ is the same for all the distributions in F. The structure of the likelihood function indicates the form that sufficient statistics take on. In particular, a necessary and sufficient condition for S to be a sufficient statistic is that the likelihood function is expressible as $L(\theta;X) = h(X,\theta)g(X)$.

From this result, known as factorization theorem, it is clear that maximum likelihood estimates are often the functions of sufficient statistics. For the multivariate normal family, the factorization theorem shows directly that the mean vector $\overline{X}$ and sample variance-covariance matrix S are sufficient statistics for the population parameters μ and Σ, respectively.

Now, we shall discuss the following theorem.

Theorem1: If $x_1, x_2, \ldots, x_n$ is a random sample from $N_p(\mu, \Sigma)$, then the maximum likelihood estimators of μ and Σ are

$$\hat{\mu} = \bar{x} \tag{11}$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right)' = \frac{1}{n}W$$

$$\hat{\Sigma} = \frac{n-1}{n}\cdot S = \frac{1}{n}W \tag{12}$$

where $W = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right)'$.

Proof: Since the x_i s are independent (because they arise from a random sample), the likelihood function (joint density) is the product of the densities of the x_i's :

$$L(\mu, \Sigma) = \prod_{i=1}^{n} f(x_i; \mu, \Sigma)$$

$$= \prod_{i=1}^{n} \frac{1}{\left(\sqrt{2\pi}\right)^p |\Sigma|^{1/2}} e^{-(x_i - \mu)'\Sigma^{-1}(x_i - \mu)/2}$$

$$= \frac{1}{\left(\sqrt{2\pi}\right)^{np} |\Sigma|^{n/2}} e^{-\sum_{i=1}^{n}(x_i - \mu)'\Sigma^{-1}(x_i - \mu)/2} \tag{13}$$

We use the notation $L(\mu, \Sigma)$ because we consider $x_1, x_2, \ldots, x_n$ to be known or available from a future sample. For the given values of $x_1, x_2, \ldots, x_n$, we seek the values of μ and Σ that maximize Eqn. (13). We first express Eqn. (13) in a form that will facilitate finding the maximum.

The scalar quantity $(x_i - \mu)'\Sigma^{-1}(x_i - \mu)$ is equal to its trace.
Hence, we have

$$\sum_{i=1}^{n}(x_i - \mu)'\Sigma^{-1}(x_i - \mu) = \sum_{i=1}^{n} tr(x_i - \mu)'\Sigma^{-1}(x_i - \mu)$$

$$= tr\left[\Sigma^{-1}\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)'\right] \tag{14}$$

$$\begin{bmatrix} \text{Here, } 1 \cdot tr(A+B) = tr(A) + tr(B) \\ 2 \cdot tr(AB) = tr(BA) \end{bmatrix}$$

Now by adding and subtracting $\bar{x}$ in the sum in the right side of Eqn. (14), we obtain

$$\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)' = \sum_{i=1}^{n}\left(x_i - \bar{x} + \bar{x} - \mu\right)\left(x_i - \bar{x} + \bar{x} - \mu\right)'$$

$$= \sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right)' + n\left(\bar{x} - \mu\right)\left(\bar{x} - \mu\right)'$$

$$= W + n\left(\bar{x} - \mu\right)\left(\bar{x} - \mu\right)' \tag{15}$$

The other two terms in the expression preceding Eqn. (15) vanish because $\sum_i \left(x_i - \bar{x}\right) = 0$. Using Eqn. (14) and Eqn. (15), we obtain

$$L(\mu, \Sigma) = \frac{1}{\left(\sqrt{2\pi}\right)^{np} |\Sigma|^{n/2}} e^{-\operatorname{tr}\Sigma^{-1}\left[w + n\left(\bar{x}-\mu\right)\left(\bar{x}-\mu\right)'\right]/2}. \tag{16}$$

Because the natural logarithm is an increasing function, the maximum of $\ln L$ will occur at the same point as the maximum of L. We prefer to work with $\ln L$ because of its simpler form for differentiation:

$$\ln L(\mu, \Sigma) = -np \ln\sqrt{2\pi} - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\operatorname{tr}\Sigma^{-1}\left[w + n\left(\bar{x}-\mu\right)\left(\bar{x}-\mu\right)'\right]$$

$$= -np \ln\sqrt{2\pi} - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}w\right) - \frac{n}{2}\left(\bar{x}-\mu\right)'\Sigma^{-1}\left(\bar{x}-\mu\right). \tag{17}$$

To find the maximum likelihood estimator for μ, we differentiate $\ln L(\mu, \Sigma)$ given in Eqn. (17) with respect to μ, and set the resulting expression equal to 0:

$$\frac{\partial \ln L\ (\mu, \Sigma)}{\partial \mu} = -0 - 0 - 0 + n\left(\Sigma^{-1}\bar{x} - \Sigma^{-1}\mu\right) = 0 \tag{18}$$

which gives

$$\hat{\mu} = \bar{x}.$$

It is clear that $\hat{\mu} = \bar{x}$ maximizes $\ln L(\mu, \Sigma)$ with respect to μ, because the last term of Eqn. (17) is ≤ 0 and the term vanishes for $\hat{\mu} = \bar{x}$.

Before differentiating $\ln L(\mu, \Sigma)$ to find $\hat{\Sigma}$, we substitute $\hat{\mu} = \bar{x}$ in Eqn. (17) and rewrite $|\Sigma|$ in term of Σ^{-1} to obtain

$$\ln L\left(\bar{\mu}, \Sigma\right) = -np \ln\sqrt{2\pi} + \frac{n}{2}\ln\left|\Sigma^{-1}\right| \frac{-1}{2}\operatorname{tr}\left(\Sigma^{-1}W\right). \tag{19}$$

We now differentiate Eqn. (19) with respect to Σ^{-1} ;

$$\frac{\partial \ln L\left(\hat{\mu}, \Sigma\right)}{\partial \Sigma^{-1}} = -0 + n\Sigma - \frac{n}{2}\operatorname{diag}(\Sigma) - W + \frac{1}{2}\operatorname{diag}(W) = 0 \tag{20}$$

From which we have

$$\hat{\Sigma} - \frac{1}{2}\operatorname{diag}\left(\hat{\Sigma}\right)\frac{1}{n}\left[W - \frac{1}{2}\operatorname{diag}(W)\right] \tag{21}$$

or $\quad \hat{\Sigma} = \frac{1}{n}W$

To see that $\hat{\Sigma} = \frac{W}{n}$ is the solution to Eqn. (21), note that for the off-diagonal elements we have $\hat{\sigma}_{jk} = w_{jk}/n, j \neq k$, and for the diagonal elements,

$\hat{\sigma}_{jj} - \hat{\sigma}_{jj}/2 = w_{jj}/n - w_{jj}/2n$ or $\hat{\sigma}_{jj}/2 = w_{jj}/2n$

[Note that we solved Eqn. (21) for Σ instead of Σ^{-1}, even though we differentiated with respect to Σ^{-1}. Otherwise we would have obtained $(W/n)^{-1}$ as the maximum likelihood estimator for Σ^{-1}. This illustrates the invariance property of maximum likelihood estimators; That is, $\left[\left(W/_n\right)^{-1}\right]^{-1} = W/n$ is the maximum likelihood estimator for $\left(\Sigma^{-1}\right)h-1=\Sigma$. Since Σ is positive definite, W is positive definite with probability1.]

Example 6: Consider a portfolio consisting of n securities, and denote the rate of return of security i by X_i. Let C_i denotes the weight of security i in the portfolio. Then, the rate of return of the portfolio is given by the linear combination,

$$R = C_1X_1 + C_2X_2 + \cdots + C_nX_n \qquad (22)$$

To see this, let x_i be the current market value of security i, and let Y_i be its future value. The rate of return of security i is given by $X_i = \frac{Y_i - x_i}{x_i}$. Let $Q_0 = \sum_{i=1}^{n} \xi_i x_i$, where ξ_i denotes the number of shares (not weight) of security i, and let Q denotes the random variable that represents the future portfolio value. Assuming that no rebalance will be made the weight C_i in the portfolio is given by,

$$C_i = \frac{\xi_i x_i}{Q_0} \qquad (23)$$

It follows that the rate of return, $R = \frac{Q-Q_o}{Q_o}$, of the portfolio is obtained as

$$R = \frac{\sum_{i=1}^{n} \xi_i (y_i - x_i)}{Q_o} = \sum_{i=1}^{n} C_i X_i \qquad (24)$$

As claimed the mean $E(R)$ and the $V(R)$ are given by,

$$E(R) = \sum_{i-1}^{n} c_i \mu_i, V(R) = \sum_{i=1}^{n}\sum_{i=1}^{n} c_i \sigma_{ij} c_j.$$

Respectively, no matter what the random variables X_i are.

Now, of interest is the potential for significant loss in the portfolio. Namely, given a probability level α, what would be the maximum loss in the portfolio ?

The maximum loss is called the value at risk (VaR) with confidence level $100\alpha\%$ and VaR is a prominent tool for risk management. More specifically, VAR with confidence level $100\alpha\%$ is defined by $z_\alpha > 0$ that satisfies

$$P\{Q - Q_0 \geq -z_\alpha\} = \alpha \qquad (25)$$

And VAR's popularity is based on aggregation of many components of market risk into the single number z_α. Fig. 1 depicts the meaning of Eqn. (25)

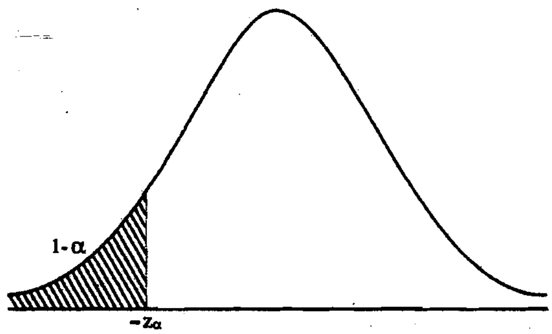


Fig. 1: The density function of $Q - Q_0$ and VaR

Since $Q - Q_o = Q_o R$, we obtain gap $P\{Q_o R \le -Z\alpha\} = 1-\alpha$ (26)

Note that the definition of VAR does not require normality. However, calculation of VAR becomes considerably simpler, if we assume that $(X_1,\cdots,X_n)$ follows an n-variate normal distribution. Then, the rate of return R, in Eqn. (24) is normally distributed with mean $E(R)=\sum_{i=1}^{n} c_i\mu_i$, and variance $V(R)=\sum_{i=1}^{n}\sum_{j=1}^{n} c_i\sigma_{ij}c_j$.

The value Z_α satisfying Eqn. (26) can be obtained using a table of the standard normal distribution. In many standard textbooks of statistics, a table of the survival probability

$$L(x)=\int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy,\ \ x>0,$$

is given. Then using the standardization $\left(Y=\frac{x-\mu}{\sigma}\right)$ and symmetry of the density function $\left(\phi(x)=\frac{1}{\sqrt{2\pi}} e^{-x^2/2}, x\in R\right)$ about 0, we can obtain the value Z_α with ease.

Namely, letting $r_\alpha = -z_\alpha/Q_0$, it follows from Eqn. (26) that

$$1-\alpha = p\left\{\frac{R-\mu}{\sigma} \le \frac{r_\alpha-\mu}{\sigma}\right\},$$

where $\mu = E(R), \sigma^2 = V(R)$ whence

$$1-\alpha = L\left(-\frac{r_\alpha-\mu}{\sigma}\right).$$

Therefore, letting x_α be such that

$L(x_\alpha)=1-\alpha$, we have $z_\alpha = Q_0(x_\alpha\sigma-\mu)$, (27)

$$\begin{aligned} -x_\alpha &= -\frac{r_\alpha-\mu}{\sigma} \\ (x_\alpha\sigma-\mu) &= -r_\alpha \\ x_\alpha\sigma-\mu &= \frac{Z_\alpha}{Q_o} \\ Z_\alpha &= Q_o(x_\alpha\sigma-\mu) \end{aligned}$$

The value z_α in Eqn. (27) is Var with confidence level $100\alpha\%$. The value x_α is the $100(1-\alpha)$ percentile of the standard normal distribution given in Table 1.

For example, if $\mu=0$, then the 99% VaR is given by $2.326\sigma Q_0$.

Note: Since the risk horizon is very short, e.g. one day or one week, in market risk management, the mean rate of return μ is often set to be zero. In this case, VAR with confidence level $100\alpha\%$ is given by $x_\alpha\sigma Q_0$.

Table 1: Percentiles of the standard normal distribution

$100(1-\alpha)$	10	5.0	1.0	0.5	0.1
x_α	1.282	1.645	2.326	2.576	3.090

Now try the following exercises:

E7) Show that $\sum_{i=1}^{n} tr(y_i-\mu)'\Sigma^{-1}(y_i-\mu) = tr\left[\Sigma^{-1}\sum_{i=1}^{n}(y_i-\mu)(y_i-\mu)'\right]$.

E8) Show that $\dfrac{\partial \ln L(\mu,\Sigma)}{\partial\mu} = n\left(\Sigma^{-1}\bar{y} - \Sigma^{-1}\mu\right)$.

E9) Let $X = \left(X_{(1)}, X_{(2)}\right)^{n}$ with $X_{(1)} = \left(X_1, \cdots X_q\right)'$ and $X_{(2)} = \left(X_{q+1}, \cdots, X_p\right)'$, and let μ be similarly partitioned as $\mu = \left(\mu_{(1)}, \mu_{(2)}\right)$, and let Σ be partitioned as $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ where Σ_{11} is a submatrix of dimension $q \times q$. Prove that if X has normal distribution with mean μ and covariance matrix Σ, and $\Sigma_{12} = \Sigma_{21} = 0$, then $X_{(1)}$ and $X_{(2)}$ are independently normally distributed with means $\mu_{(1)}, \mu_{(2)}$ and covariance matrices Σ_{11}, Σ_{12}, respectively.

E10) Show that $\sum_{j=1}^{n} (x_j - \bar{x})(\bar{x} - \mu)'$ and $\sum_{j=1}^{n} (\bar{x} - \mu)(x_j - \bar{x})'$ are both $p \times p$ matrices of zeros. Here $x'_j = [x_{j1}, x_{j2}, \ldots, x_{jp}]$, $j = 1, 2, \ldots, n$ and $\bar{x} = \dfrac{1}{n}\sum_{j=1}^{n} x_j$.

Now let us sum up this unit.

19.4 SUMMARY

In this unit, we have covered the following points.

1. A multivariate normal distribution is a generalization of the one-dimensional normal distribution to higher dimensions. It is also closely related to matrix normal distribution.

2. Estimation of parameters of multivariate normal distribution requires maximum likelihood estimation (MLE), which is a popular statistical methods used for fitting a mathematical model to some data. Maximum likelihood estimation gives a unique and easy to determine solution in the case of normal distribution and many other problems. That is, it has widespread applications in various fields, including structural equation modeling, psychometrics and econometrics, data modeling in nuclear and particle physics, path-choice modeling in transport networks, etc.

19.5 SOLUTIONS/ANSWERS

E1) $$\begin{bmatrix} X_1 - X_2 \\ X_2 - X_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = AX \quad \text{(suppose)}$$

the distribution of AX is multivariate normal with mean

$$A\mu = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_2 - \mu_3 \end{bmatrix}$$

And covariance matrix

$$A\Sigma A' = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} - 2\sigma_{12} + \sigma_{22} & \sigma_{12} + \sigma_{23} - \sigma_{22} - \sigma_{13} \\ \sigma_{12} + \sigma_{23} - \sigma_{22} - \sigma_{13} & \sigma_{22} - 2\sigma_{23} + \sigma_{33} \end{bmatrix}$$

Alternatively, the mean vector $A\mu$ and covariance matrix $A\Sigma A'$ may be verified by direct calculation of the means and covariances of the two random variables $Y_1 = X_1 - X_2$ and $Y_2 = X_2 - X_3$.

E2) We set $X_1 = \begin{bmatrix} X_2 \\ X_4 \end{bmatrix}, \mu_1 = \begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}, \Sigma_{11} = \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix}$

And note that with this assignment $X, \mu,$ and Σ can respectively be rearranged and partitioned as

$$X = \begin{bmatrix} X_2 \\ X_4 \\ \cdots \\ X_1 \\ X_3 \\ X_5 \end{bmatrix}, \mu = \begin{bmatrix} \mu_2 \\ \mu_4 \\ \cdots \\ \mu_1 \\ \mu_3 \\ \mu_5 \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_{22} & \sigma_{24} & \vdots & \sigma_{12} & \sigma_{23} & \sigma_{25} \\ \sigma_{24} & \sigma_{44} & \vdots & \sigma_{14} & \sigma_{34} & \sigma_{45} \\ \cdots & \cdots & \vdots & \cdots & \cdots & \cdots \\ \sigma_{12} & \sigma_{14} & \vdots & \sigma_{11} & \sigma_{13} & \sigma_{15} \\ \sigma_{23} & \sigma_{34} & \vdots & \sigma_{13} & \sigma_{33} & \sigma_{35} \\ \sigma_{25} & \sigma_{45} & \vdots & \sigma_{15} & \sigma_{35} & \sigma_{55} \end{bmatrix}$$

or,

$$X = \begin{bmatrix} \underset{(2\times1)}{X_1} \\ \cdots \\ \underset{(3\times1)}{X_2} \end{bmatrix}, \mu = \begin{bmatrix} \underset{(2\times1)}{\mu_1} \\ \cdots \\ \underset{(3\times1)}{\mu_2} \end{bmatrix}, \Sigma = \begin{bmatrix} \underset{(2\times2)}{\Sigma_{11}} & \vdots & \underset{(2\times3)}{\Sigma_{12}} \\ \cdots & \vdots & \cdots \\ \underset{(3\times2)}{\Sigma_{21}} & \vdots & \underset{(3\times3)}{\Sigma_{22}} \end{bmatrix}$$

Thus, from the distribution

$$N_2(\mu_1, \Sigma_{11}) = N_2\left(\begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}, \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix}\right)$$

It is clear from this example that the normal distribution for any subset can be expressed by simply selecting the appropriate means and covariances from the original μ & Σ.

E3) Since Σ is diagonal matrix of order p, let us prove the same for p = 2.

for $p = 2$, $\Sigma = \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix}$

here, $\sigma_{12} = \sigma_{21} = 0$ as Σ is a diagonal matrix.

The density function for bivariate normal distribution can be given by:

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}\left(1-\rho_{12}^2\right)}}$$

$$\text{x } \exp\left\{-\frac{1}{2\left(1-\rho_{12}^2\right)}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]\right\}$$

since $\sigma_{12} = 0 \therefore \rho_{12} =$ correlation coefficient $= 0$

this gives,

$$f(x_1,x_2)=\frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}}} x \exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2+\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2\right]\right\}$$

$$=\left[\frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}}}\cdot\exp\left\{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2\right\}\right]\cdot\left[\frac{1}{\sqrt{2\pi}\sqrt{\sigma_{22}}}\cdot\exp\left\{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2\right\}\right]$$

$f(x_1,x_2)=f(x_1).f(x_2)$

i.e. if the random variables X_1 and X_2 are uncorrelated, so that $\rho_{12}=0$, then the joint density are the product of two univariate normal density and X_1, X_2 are independent.

This result is true in general.

E4) $\mu=\begin{bmatrix}2\\3\\1\end{bmatrix}$.

$e_1=\begin{bmatrix}-3\\0\\3\end{bmatrix}$, $e_2=\begin{bmatrix}0\\1\\-1\end{bmatrix}$ and $e_3=\begin{bmatrix}-3\\1\\2\end{bmatrix}$

$s_n=\begin{bmatrix}6 & -1 & 5\\-1 & 2/3 & -1/3\\5 & -1/3 & 14/3\end{bmatrix}$.

E5) a) Since x_1 and x_2 have covariance $\sigma_{12}-2$, they are not independent.

b) Since x_1 and x_3 have covariance $\sigma_{13}=0$, they are independent.

c) Partitioning x and Σ in pairs (x_1, x_2) and x_3, we get

$$x=\begin{bmatrix}x_1\\x_2\\\cdots\\x_3\end{bmatrix}, \Sigma=\begin{bmatrix}1 & -2 & \vdots & 0\\-2 & 5 & \vdots & 0\\\cdots & \cdots & \vdots & \cdots\\0 & 0 & \vdots & 2\end{bmatrix}=\begin{bmatrix}\Sigma_{11} & \vdots & \Sigma_{12}\\\cdots & \vdots & \cdots\\\Sigma_{21} & \vdots & \Sigma_{22}\end{bmatrix}.$$

Now here $\begin{bmatrix}x_1\\x_2\end{bmatrix}$ and x_3 have covariance matrix $\Sigma_{12}=\begin{bmatrix}0\\0\end{bmatrix}$, therefore (x_1, x_2) and x_3 are independent.

d) Similarly, partitioning x and Σ in $\frac{x_1+x_2}{2}$ and x_3,

$$x=\begin{bmatrix}\frac{x_1+x_2}{2}\\\cdots\cdots\cdots\\x_3\end{bmatrix}$$

Further, $\begin{bmatrix}\frac{x_1+x_2}{2}\\x_3\end{bmatrix}=\begin{bmatrix}1/2 & 1/2 & 0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$

Thus, $A = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Now,

$$A\Sigma A' = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1/2 & 0 \\ 1/2 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1/2 & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & 2 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \vdots & \Sigma_{12} \\ \cdots & \vdots & \cdots \\ \Sigma_{21} & \vdots & \Sigma_{22} \end{bmatrix}$$

Since the covariance matrix $\Sigma_{12} = 0$ (which is the covariance between $\frac{x_1 + x_2}{2}$ and x_3), therefore, $\frac{x_1 + x_2}{2}$ and x_3 are independent.

E7) $\sum_{i=1}^{n} \text{tr}(y_i - \mu)' \Sigma^{-1}(y_i - \mu) = \text{tr}\,\Sigma^{-1} \sum_{i=1}^{n} (y_i - \mu)(y_i - \mu)'$

It follows from the fact that "trace of a sum of matrices is equal to the sum of the traces of the matrices".

Also, $x'Ax = \text{tr}(x'Ax) = \text{tr}(Axx')$ where A is a $k \times k$ symmetric matrix and x be a $k \times 1$ vector.

E8) Since we know that

$$\ln L(\mu, \Sigma) = -np \ln\sqrt{2\pi} - \frac{n}{2} |\, n \,|\, |\Sigma| - \frac{1}{2}\text{tr}(\Sigma^{-1}w) - \frac{n}{2}(\bar{y} - \mu)'\Sigma^{-1}(\bar{y} - \mu)$$

$$\Rightarrow \frac{\partial \ln L(\mu, \Sigma)}{\partial \mu} = 0 - 0 - 0 - n(-\Sigma^{-1}(\bar{y} - \mu))$$

$$= n(\Sigma^{-1}\bar{y} - \Sigma^{-1}\mu)$$

(since $\frac{n}{2}(\bar{y} - \mu)' \Sigma^{-1}(\bar{y} - \mu) = \frac{n}{2}(\bar{y} - \mu)^2 \Sigma^{-1}$ and

$\frac{\partial}{\partial \mu}\left[\frac{n}{2}\Sigma^{-1}(\bar{y} - \mu)^2\right] = -n\Sigma^{-1}(\bar{y} - \mu)$.

E9) Given, $x = \begin{bmatrix} x_{(1)} \\ x_{(2)} \end{bmatrix}$ $\mu = \begin{bmatrix} \mu_{(1)} \\ \mu_{(2)} \end{bmatrix}$

$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \Sigma^{-1} = \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix}$ (Since $\Sigma_{12} = \Sigma_{21} = 0$)

Also, x is $N_p(\mu, \Sigma)$ with $|\Sigma| \neq 0$,

Now, $(x - \mu)'\Sigma^{-1}(x - \mu) = [(x_{(1)} - \mu_{(1)})'(x_{(2)} - \mu_{(2)})']$

$$\begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} x_{(1)} & \mu_{(1)} \\ x_{(2)} & \mu_{(2)} \end{bmatrix}$$

$$= (x_{(1)} - \mu_{(1)})' \Sigma_{11}^{-1}(x_{(1)} - \mu_{(1)}) + (x_{(2)} - \mu_{(2)})' \Sigma_{22}^{-1}(x_{(2)} - \mu_{(2)})$$

i.e., $x_{(1)}$ and $x_{(2)}$ are independently normally distributed with mean $\mu_{(1)}, \mu_{(2)}$ respectively and covariances matrices Σ_{11} and Σ_{22} respectively.

E10) Let us take, $\sum_{j=1}^{n} (x_j - \bar{x})(\bar{x} - \mu)'$

$$= \sum_{j=1}^{n} x_j \bar{x} - \sum_{j=1}^{n} x_j \mu - \sum_{j=1}^{n} \bar{x}^2 + \sum_{j=1}^{n} \bar{x}\,\mu$$

$$= n\,\bar{x}^2 - n\,\bar{x}\mu - n\,\bar{x}^2 + n\,\bar{x}\,\mu = 0$$

Similarly, other can be solved.

19.6 PRACTICAL ASSIGNMENT

Session 4

1. Write a programme is 'C' language to find the ML estimation of mean (μ) and variance (σ^2).

UNIT 20 INFERENCE-II

Structure **Page No**

20.1 INTRODUCTION

We shall begin the unit by discussing the plausibility of μ_0 as a value for a normal population mean in Sec. 20.2. Multivariate hypothesis testing differs from univariate testing in many ways. For example, the number of possible hypotheses is very large and in many cases, there are several alternative test statistics from which to choose. The number of possible hypotheses is large because of the large number of parameters. For example, the p-variate normal distribution has $\frac{1}{2}p(p+3)$ parameters, for each of which we could specify a hypothesis, as well as for subsets or functions of the parameters. There are many advantages to testing p variables in a single multivariate test rather than in p separate univariate tests. These advantages include preserving the α – level, testing with greater power, and determining the contribution of each variable in the presence of the other variables. The multivariate tests make allowance for the intercorrelations among the variables, which accounts in part for the increase in power. At this point, we shall concentrate on inferences about a population mean vector and its components through Hotelling's T^2 test in section 20.3. At the end of this unit we shall discuss the sample distribution of $\overline{X}$ and S in section 20.4.

Objectives

After studying this unit, you should be able to:

- make on inferences about a population mean vector and covariance matrix;
- understand the correspondence between univariate and multivariate test statistic;
- compute Hotelling's T^2;
- test the hypothesis with T^2;
- define the distribution of mean and sample covariance;
- decide whether a process is in control or not.

20.2 INFERENCES ABOUT μ_0 AS A VALUE FOR NORMAL POPULATION MEAN

Let us start by recalling the univariate theory for determining whether a specific value μ_0 is a plausible value for the population mean μ. From the point of view of hypothesis testing, this problem can be formulated as a test for comparing hypotheses

$$H_0 : \mu = \mu_0 \text{ and } H_1 : \mu \neq \mu_0$$

Here H_0 is the null hypothesis and H_1 is the (two-sided) alternative hypothesis. If $X_1, X_2, \cdots, X_n$ denote a random sample from a normal population, the appropriate test statistic to be used is given by

$$t = \frac{\overline{X} - \mu_0}{S/\sqrt{n}},$$

where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and

$$S^2 = \frac{1}{n-1}\sum_{j=1}^{m} (X_j - \overline{X})^2$$

This test statistic has a student's t distribution with $n-1$ degrees of freedom (d.f.). We reject H_0, at some specified level of significance, which implies that μ_0 cannot be a plausible value of μ. Rejecting H_0 when $|t|$ is large is equivalent to rejecting H_0 if its square,

$$t^2 = \frac{\left(\bar{X} - \mu_0\right)^2}{S^2/n} = n\left(\bar{X} - \mu_0\right)\left(S^2\right)^{-1}\left(\bar{X} - \mu_0\right) \tag{1}$$

is large. The variable t^2 in Eqn. (1) is the square of the distance from the sample mean $\overline{X}$ to the test value μ_0. The units of distance are expressed in terms of $S/\sqrt{n}$, or estimated standard deviations of $\overline{X}$. Once $\overline{X}$ and S^2 are observed, the test is given by Reject H_0 in favour of H_1, at significance level α, if

$$n\left(\bar{X} - \mu_0\right)\left(S^2\right)^{-1}\left(\bar{X} - \mu_0\right) > t_{n-1}^{\;2}(\alpha/2) \tag{2}$$

where $t_{n-1}(\alpha/2)$ denotes the upper $100(\alpha/2)^{th}$ percentile of the t-distribution with $n-1$ d.f. If H_0 is not rejected, we conclude that μ_0 is a plausible value for the normal population mean. For a given data set, there are always a set of plausible values of μ that are consistent with the given data. From the well-known correspondence between acceptance regions for tests of $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ and confidence intervals for μ, we have

$\{\text{Do not reject } H_0 : \mu = \mu_0 \text{ at level } \alpha\}$ or

$$\left|\frac{\bar{X} - \mu_0}{S/\sqrt{n}}\right| \leq t_{n-1}(\alpha/2)$$

which is equivalent to

$$\left\{\mu_0 \text{ lies in the } 100(1-\alpha)\}\% \text{ confidence interval } \bar{x} \pm t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}}\right\}. \tag{3}$$

The confidence interval consists of all those values μ_0 that would not be rejected by the level α test of $H_0 : \mu = \mu_0$. Before the sample is selected, the $100(1-\alpha)\%$ confidence interval in Eqn. (3) is a random interval because the endpoints depend upon the random variables $\overline{X}$ and S. The probability that the interval contains μ is $1-\alpha$; among large numbers of such independent intervals, approximately $100(1-\alpha)\%$ of them will contain μ.

Consider now the problem of determining whether a given $p \times 1$ vector μ_0 is a plausible value for the mean of a multivariate normal distribution $N_p(\mu, \Sigma)$, Σ unknown. We shall proceed by analogy to the univariate development just presented. A natural generalization of the squared distance in Eqn. (1) is its multivariate analog

$$T^2 = \left(\bar{X} - \mu_0\right)'\left(\frac{1}{n}S\right)^{-1}\left(\bar{X} - \mu_0\right)$$

$$T^2 = n\,(\bar{X} - \mu_0)'\,S^{-1}\,(\bar{X} - \mu_0) \tag{4}$$

where $\underset{p\times 1}{\bar{X}} = \frac{1}{n}\sum_{j=1}^{n} X_j$,

$$\underset{(p\times p)}{S} = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \bar{X}\right)\left(X_j - \bar{X}\right)'$$

and $\underset{p\times 1}{\mu_0} = \begin{bmatrix} \mu_{10} \\ \mu_{20} \\ \vdots \\ \mu_{10} \end{bmatrix}$.

The statistic T^2 is called Hotelling's T^2 in honor of Harold Hotelling, a pioneer in multivariate analysis, who first obtained its sampling distribution. Here $(1/n)S$ is the estimated covariance matrix of $\bar{X}$.

If the observed statistical distance T^2 is too large – that is, if $\bar{X}$ is "too far" from μ_0 - the hypothesis $H_0 : \mu = \mu_0$ is rejected. It turns out that special tables of T^2 percentage points are not required for formal tests of hypothesis. This is true because T^2 is distributed as $\frac{(n-1)p}{n-p}F_{p,n-p}$, (5)

where $F_{p,n-p}$ denotes a random variables with an F-distribution with p and $n-p$ d.f.

Note: The square of a univariate t has an $F_{1,\,n-1}$ distribution. To summarize, we have the following:

Let $X_1, X_2, \cdots, X_n$ be a random sample from an $N_p(\mu, \Sigma)$ population. Then with

$\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$ and $S = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \bar{X}\right)\left(X_j - \bar{X}\right)'$,

$$\alpha = p\left[T^2 > \frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)\right]$$

$$= p\left[n\left(\bar{x} - \mu\right)' S^{-1}\left(\bar{x} - \mu\right) > \frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)\right] \tag{6}$$

Whatever the true μ and Σ be. Here $F_{p,n-p}(\alpha)$ is the upper (100α)th percentile of the $F_{p,n-p}$ distribution.

Statement given in Eqn. (6) leads immediately to a test (in the multivariate case, a simple function of T^2 also has an F-distribution) of the hypothesis $H_0 : \mu = \mu_0$ against $H_0 : \mu \neq \mu_0$. At the α level of significance, we reject H_0 in favour of H_1 if the observed

$$T^2 = n\left(\bar{X} - \mu_0\right)' S^{-1}\left(\bar{X} - \mu_0\right) > \frac{(n-1)p}{n-p}F_{p,n-p}(\alpha) \tag{7}$$

In the following section, we shall discuss Hotelling's T^2-Distribution.

20.3 HOTELLING'S T^2-DISTRIBUTION

The formal definition of a T^2 random variable is similar to the formal definition of the univariate t^2 random variable. Let Z be distributed as the multivariate normal $N_p(0,\Sigma)$ and W be distributed as Wishart $W_p(\nu,\Sigma)$, with Z and W independent. Then the T^2 random variable with dimension p and degrees of freedom ν is defined as

$$T^2 = Z'\left(\frac{W}{\nu}\right)^{-1} Z \tag{8}$$

The distribution of Hotelling's T^2 can be derived from this definition.

Define $\bar{V} = \sqrt{n}\left(\bar{X} - \mu_0\right)$ and $W = (n-1)S$.

Then $\bar{V}$ is $N_p(0,\Sigma)$ if $\mu = \mu_0$, W is $W_p(n-1,\Sigma)$, and $\bar{V}$ and W are independent. Hence by using Eqn. (8) we have

$$T^2 = \bar{V}\left(\frac{W}{n-1}\right)^{-1}\bar{V}$$

$$= \left[\sqrt{n}\,(\bar{X}-\mu_0)\right]\left[\frac{(n-1)S^{-1}}{n-1}\right]^{-1}\left[\sqrt{n}\,(\bar{X}-\mu_0)\right]$$

$$= n(\bar{X}-\mu_0)'\,S^{-1}(\bar{X}-\mu_0),$$

which is given in Eqn. (4).

We now consider the case in which the hypothesis $H_0 : \mu = \mu_0$ is p-dimensional. In order to test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$, we assume that a random sample $X_1, X_2, \cdots, X_n$ is available from $N_p(\mu,\Sigma)$, with Σ unknown. By analogy with the univariate t-statistic' and the Z^2-Statistic and also, from Theorem 1in Unit 3, we use the sample mean vector $\bar{X}$ and the sample covariance matrix S to construct the test statistic,

$$T^2 = n\left(\bar{X}-\mu_0\right)' S^{-1}\left(\bar{X}-\mu_0\right) \tag{9}$$

which is the standardized distance from $\bar{X}$ to μ_0.

If H_0 is true and if sampling is from $N_p(\mu,\Sigma)$, then T^2 in Eqn. (9) has Hotelling's T^2-distribution with dimension p and degrees of freedom $(n-1)$. We reject H_0 if

$T^2 \geq T^2_{\alpha,p,n-1} = N_p(0,\Sigma)'\left[\frac{1}{n-1}W_{p,\,n-1}(\Sigma)\right]^{-1} N_p(0,\Sigma)$ and accept H_0 otherwise.

For $p=1$, the T^2 statistic in Eqn. (8) reduces to the square of the univariate t as shown:

$$T^2 = n\left(\bar{x}-\mu_0\right)\left(S^2\right)^{-1}\left(\bar{x}-\mu_0\right) = \frac{n\left(\bar{x}-\mu_0\right)^2}{S^2} = t^2.$$

Another link between T^2 and the univariate t is that in cases where there is an analogous t-test, the degrees of freedom for the T^2-test will be same as for the univariate t-test. Thus the one-sample T^2-test has $n-1$ degrees of freedom, and the two-sample T^2-test has $n_1 + n_2 - 2$ degrees of freedom. The two-sample T^2-test is defined as:

if $X_{11}, X_{12}, \ldots, X_{1n1}$, is random sample of size n_1 from $N_p(\mu_1, \Sigma)$ and $X_{21}, X_{22}, \ldots, X_{2n2}$ is an independent random sample of size n_2 from $N_p(\mu_2, \Sigma)$,

then $T^2 = N_p(0,\Sigma)' \left[\dfrac{W_{n_1+n_2-2}(\Sigma)}{n_1+n_2-2}\right]^{-1} N_p(0, \Sigma)$.

A key assumption in the T^2-distribution is the independence of $\overline{X}$ & S; this assumption holds when sampling is from a multivariate normal population. Also if $\overline{X}$ & S are based on a random sample $X_1, X_2, \cdots, X_n$ from $N_p(\mu, \Sigma)$, then $\overline{X}$ and S are independent). Usually $\overline{X}$ & S are obtained from the same sample, although this is not necessary. As long as S is an unbiased estimator of Σ and is independent of $\overline{X}$, the estimator S could come wholly or partly from another sample. For example, suppose $\bar{X}_1$ and S_1 arise from a sample of size n_1 from $N_p(\mu_1, \Sigma)$ and that a supplementary estimate S_2 is available from a sample of size n_2 from $N_p(\mu_2, \Sigma)$, where the two distributions have a common covariance matrix Σ and the two sample are independent. We estimate Σ by the pooled estimator

$S_{pl} = [(n_1-1)S_1 + (n_2-1)S_2]/(n_1+n_2-2)$, which has n_1+n_2-2 degrees of freedom. Then a T^2-statistic could be based on S_1, or S_2, or S_{pl}:

$$n_1\left(\bar{X}_1-\mu_1\right)' \; S_1^{-1}\left(\bar{X}_1-\mu_1\right) \quad \text{is} \quad T_p^2, n_1-1,$$

$$n_1\left(\bar{X}_1-\mu_1\right)' \; S_2^{-1}\left(\bar{X}_1-\mu_1\right) \quad \text{is} \quad T_p^2, n_2-1;$$

$$n_1\left(\bar{X}_1-\mu_1\right)' \; S_{pl}^{-1}\left(\bar{X}_1-\mu_1\right) \quad \text{is} \quad T_p^2, n_1+n_2-2.$$

The coefficient of the quadratic form in all three cases is n_1 because each of S_1/n_1, S_2/n_1 and S_{pl}/n_1 estimates $\dfrac{\Sigma}{n_1}$, the covariance matrix of $\overline{X}_1$. Thus, the leading coefficient in these T^2-Statistic is the sample size for $\overline{X}_1$.

Some important properties of the T^2-test statistic for testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ are as follows:

1. A T^2-Statistic is invariant to the transformations of the form $Z_i = AX_i + b$, where A is a non-singular matrix that is $T_z^2 = T_x^2$. Invariance of this type, often referred to as affine invariance includes changes in scale, as, for example, from inches to centimeters.
2. The T^2-Statistic is the uniformly most powerful (UMP) invariant test. UMP test is defined as follows: In statistical hypothesis testing, a UMP test is a hypothesis test which has the greatest power $1-\beta$ among all possible tests of a given size α.
3. The T^2-Statistic is sensitive to departures from normality.
4. The T^2-Statistic is the likelihood ratio test statistic.
5. The T^2-Statistic is the union-intersection test statistic.

In the univariate case $(p=1)$, the t-statistic can be adapted to serve for a one sided alternative hypothesis such as $H_1 : \mu > \mu_0$. The resulting one-tailed t-test has some optimal properties. In the multivariate case, however, the T^2-Statistic cannot be similarly adapted to have optimal properties for a one-sided alternative. For $p > 1$, the one-sided alternative $H_1 : \mu > \mu_0$ can be defined to mean that $\mu_j > \mu_{0j}$ for all $j = 1, 2, \cdots, p$. Kariya and Cohen (1992) showed that for this case there is no scale invariant test statistic with satisfactory properties. The T^2-test is invariant but cannot be recommended for obvious reasons – it would reject H_0 when we want to accept it, namely, when same or all $\overline{X}_j$ are considerably less than the corresponding μ_{0j}.

Now again consider

$$T^2 = n\left(\bar{X} - \mu_0\right)' S^{-1}\left(\bar{X} - \mu_0\right)$$

We can write

$$T^2 = \sqrt{n}\left(\bar{X} - \mu_0\right)' \left(\frac{\sum_{j=1}^{n}\left(X_j - \bar{X}\right)\left(X_j - \bar{X}\right)'}{n-1}\right)^{-1} \sqrt{n}\left(\bar{X} - \mu_0\right)$$

which combines a normal, $N_p(0, \Sigma)$, random vector and a Wishart, $W_{p,n-1}(\Sigma)$, random matrix in the form

$$T^2_{p,n-1} = \begin{pmatrix}\text{multivariate}\\ \text{normal random}\\ \text{vector}\end{pmatrix}' \begin{pmatrix}\dfrac{\text{Wishart random matrix}}{\text{d.f.}}\end{pmatrix}^{-1} \begin{pmatrix}\text{multivariate}\\ \text{normal random}\\ \text{vector}\end{pmatrix} \tag{10}$$

$$= N_p(0, \Sigma)' \left[\frac{1}{n-1} W_{p,n-1}(\Sigma)\right]^{-1} N_p(0, \Sigma).$$

This is analogous to

$$t^2 = \sqrt{n}\left(\bar{X} - \mu_0\right)\left(S^2\right)^{-1} \sqrt{n}\left(\bar{X} - \mu_0\right)$$

or,

$$t^2_{n-1} = \begin{pmatrix}\text{Normal}\\ \text{random}\\ \text{variable}\end{pmatrix} \begin{pmatrix}\dfrac{\text{Chi-square random variable}}{\text{d.f.}}\end{pmatrix}^{-1} \begin{pmatrix}\text{Normal}\\ \text{random}\\ \text{variable}\end{pmatrix}$$

for the univariate case. Since the multivariate normal and Wishart random variables are independently distributed, their joint density function is the product of the marginal normal and Wishart distributions. Using calculus, the distribution of T^2 can be derived from the joint distribution given in Eqn.(10) of multivariate normal and Wishart distributions.

Example 1 (Evaluating T^2): Let the data matrix for a random sample of size $n = 3$ from a bivariate normal population be

$$X = \begin{bmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \end{bmatrix}$$

Evaluate the observed T^2 for $\mu_0' = [9, 5]$. What is the sampling distribution of T^2 in this case?

Solution: We find

$$\overline{X} = \begin{bmatrix} \overline{X}_1 \\ \overline{X}_2 \end{bmatrix} = \begin{bmatrix} \frac{6+10+8}{3} \\ \frac{9+6+3}{3} \end{bmatrix} = \begin{bmatrix} 8 \\ 6 \end{bmatrix}$$

And

$$S_{11} = \frac{(6-8)^2 + (10-8)^2 + (8-8)^2}{2} = 4$$

$$S_{12} = \frac{(6-8)(9-6) + (10-8)(6-6) + (8-8)(3-6)}{2} = -3$$

$$S_{22} = \frac{(9-6)^2 + (6-6)^2 + (3-6)^2}{2} = 9$$

$$S = \begin{bmatrix} 4 & -3 \\ -3 & 9 \end{bmatrix}$$

Thus $S^{-1} = \frac{1}{(4)(9)-(-3)(-3)} \begin{bmatrix} 9 & 3 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{9} \\ \frac{1}{9} & \frac{4}{27} \end{bmatrix}$

From, $T^2 = n\left(\bar{X} - \mu_0\right)' S^{-1} \left(\bar{X} - \mu_0\right)$

We have,

$$T^2 = 3[8-9, 6-5] \begin{bmatrix} \frac{1}{3} & \frac{1}{9} \\ \frac{1}{9} & \frac{4}{27} \end{bmatrix} \begin{bmatrix} 8 & - & 9 \\ 6 & - & 5 \end{bmatrix}$$

$$T^2 = 3[-1 \quad 1] \begin{bmatrix} -\frac{2}{9} \\ \frac{1}{27} \end{bmatrix} = \frac{7}{9}$$

Before the sample is selected, T^2 has the distribution of a

$$\frac{(3-1)^2}{(3-2)} F_{2,3-2} = 4 F_{2,1} \text{ random variable.}$$

Example 2 (Testing a multivariate mean vector with T^2): Perspiration from 20 healthy females was analyzed. Three components $X_1 = \text{Sweat rate}$, $X_2 = \text{Sodium content, and } X_3 = \text{potassium content}$, were measured, and the results, which we call the sweat data, are presented in Table 1. Test the hypothesis $H_0 : \mu' = [4, 50, 10]$ against $H_1 : \mu' \neq [4, 50, 10]$ at level of significance $\alpha = 0.10$.

Table 1-Sweat Data

Individual	X_1	X_2	X_3
1	3.7	48.5	9.3
2	5.7	65.1	8.0
3	3.8	47.2	10.9
4	3.2	53.2	12.0
5	3.1	55.5	9.7
6	4.6	36.1	7.9
7	2.4	24.8	14.0
8	7.2	33.1	7.6

9	6.7	47.4	8.5
10	5.4	54.1	11.3
11	3.9	36.9	12.7
12	4.5	58.8	12.3
13	3.5	27.8	9.8
14	4.5	40.2	8.4
15	1.5	13.5	10.1
16	8.5	56.4	7.1
17	4.5	71.6	8.2
18	6.5	52.8	10.9
19	4.1	44.1	11.2
20	5.5	40.9	9.4

Solution: From the data given in Table 1

$$\bar{X} = \begin{bmatrix} 4.640 \\ 45.400 \\ 9.965 \end{bmatrix} \quad S = \begin{bmatrix} 2.879 & 10.010 & -1.810 \\ 10.010 & 199.788 & -5.640 \\ -1.810 & -5.640 & 3.628 \end{bmatrix}$$

And

$$S^{-1} \begin{bmatrix} 0.586 & -0.022 & 0.258 \\ -0.022 & 0.006 & -0.002 \\ 0.258 & -0.002 & 0.402 \end{bmatrix}$$

We evaluate

$$T^2 = 20[4.640-4, 45.400-50, 9.965-10]S^{-1}\begin{pmatrix} 4.640-4 \\ 45.400-50 \\ 9.965-10 \end{pmatrix}$$

$$T^2 = 9.74$$

Comparing the observed $T^2 = 9.74$ with the critical value

$$\frac{(n-1)p}{(n-p)}F_{p,n-p}(.10) = \frac{(19)(3)}{17}F_{3,17}(.10)$$

$$= 3.353(2.44) \; [F_{3,17}(0.10) = 2.44 \text{ from the table given in the Appendix}]$$

$$= 8.18$$

We see that $T^2 = 9.74 > 8.18$, and consequently, we reject H_0 at the 10% level of significance.

Try the following exercises:

E1) (a) Evaluate T^2, for testing $H_0 : \mu' = [7, 11]$, using the data

$$X = \begin{bmatrix} 2 & 12 \\ 8 & 9 \\ 6 & 9 \\ 8 & 10 \end{bmatrix}$$

(b) Specify the distribution of T^2 for the situation in (a).

E2) For $n = 42$ pairs of observations from the bivariate normal distribution, the quantities $\bar{X}$, S and S^{-1} are given as follows:

$$\bar{X} = \begin{bmatrix} 0.564 \\ 0.603 \end{bmatrix} S = \begin{bmatrix} 0.0144 & 0.0117 \\ 0.0117 & 0.0146 \end{bmatrix} S^{-1} = \begin{bmatrix} 203.018 & -163.391 \\ -163.391 & 200.228 \end{bmatrix}$$

Conduct a test of the null hypothesis $H_0 : \mu' = [0.55 \quad 0.60]$ at the $\alpha = 0.05$ level of significance.

E3) Using the data in E1(a), verify that T^2 remains unchanged if each observation $X_j, j = 1,2,3$ is replaced by CX_j, where

$$C = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Let us now discuss the sampling distribution of $\overline{X}$ and S.

20.4 THE SAMPLING DISTRIBUTION OF $\overline{X}$ AND S

The tentative assumption that $X_1, X_2, \cdots, X_n$ constitute a random sample from a normal population with mean μ and covariance Σ completely determines the sampling distributions of $\overline{X}$ and S. Here we present the result on the sampling distributions of $\overline{X}$ and S by drawing a parallel with the familiar univariate conclusions.

In the univariate case $(p = 1)$, we know that $\overline{X}$ is normal with mean $\mu = (\text{population mean})$ and variance.

$$\frac{1}{n}\sigma^2 = \frac{\text{Population Variance}}{\text{Sample Size}}$$

The result for the multivariate case $(p \geq 2)$ is analogous in that $\overline{X}$ has a multivariate normal distribution with mean μ and covariance matrix $(1/n)\Sigma$. Again for p = 1 for the sample variance, recall that $(n-1)S^2 = \sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2$ is distributed as σ^2 times a chi-square variable having $n-1$ degrees of freedom (d.f.). In turn, this chi-square is the distribution of a sum of squares of independent standard normal random variables. That is, $(n-1)S^2$ is distributed as $\sigma^2\left(Z_1^2 + \cdots + Z^2_{n-1}\right) = (\sigma Z_1)^2 + \cdots + (\sigma Z_{n-1})^2$. The individual terms σZ_i are independently distributed as $N\left(0, \sigma^2\right)$.

It is this latter form that is suitably generalized to the basic sampling distribution for the sample covariance matrix.

The sampling distribution of the sample covariance matrix is called the Wishart distribution, after its discoverer; it is defined as the sum of independent products of multivariate normal random vectors. Specifically,

$$\begin{aligned} W_m(.1\Sigma) &= \text{Wishart distribution with m d.f.} \\ &= \text{Distribution of} \sum_{j=1}^{m} Z_j Z_j' \end{aligned}$$

where Z_j are each independently distributed as $N_p(0, \Sigma)$.

We summarize the sampling distribution results as follows:

Let $X_1, \cdots, X_n$ be a random sample of size n from a p-variate normal distribution with mean μ and covariance matrix Σ. Then

1. $\overline{X}$ is distributed as $N_p\left(\mu, \left(\frac{1}{n}\right)\Sigma\right)$.
2. $(n-1)S$ is distributed as a Wishart random matrix with $n-1$ d.f..

3. $\bar{X}$ and S are independent.

Because Σ is unknown, the distribution of $\bar{X}$ cannot be used directly to make inferences about μ. However, S provides independent information about Σ, and the distribution of S does not depend on μ. This allows us to construct a statistic for making inferences about μ.

Properties of the Wishart Distribution

1. If A_1 is distributed as $W_{m1}(A_1/\Sigma)$ independently of A_2, which is distributed as $W_{m2}(A_2/\Sigma)$, then $A_1 + A_2$ is distributed as $W_{m1+m2}(A_1 + A_2/\Sigma)$. That is, the degrees of freedom are added.
2. If A is distributed as $W_m(A/\Sigma)$, then CAC' is distributed as $W_m(CAC'/CEC')$.

Let us discuss the Large-Sample Behaviour of $\bar{X}$ and S.

Suppose the quantity X is determined by a large number of independent causes $V_1, V_2, \cdots, V_n$, where the random variables V_i representing the causes have approximately the same variability. If X is the sum

$$X = V_1 + V_2 + \cdots + V_n$$

Then the central limit theorem applies, and we conclude that X has a distribution that is nearly normal. This is true for virtually any parent distribution of the V_i's, provided that n is large enough.

The univariate central limit theorem also tells us that the sampling distribution of the sample mean, $\bar{X}$ for a large sample size is nearly normal, whatever the form of the underlying population distribution.

It turns out that certain multivariate statistics, like $\bar{X}$ and S, has large-sample properties analogous to their univariate counterparts. As the sample size is increased without bound, certain regularities govern the sampling variation in $\bar{X}$ and S, irrespective of the form of the parent population. Therefore, the conclusions presented in this section do not require multivariate normal populations. The only requirement are that the parent population, whatever its form, have a mean μ and a finite covariance Σ.

Example 3 (Multivariate Quality Control Charts for Monitoring a Sample of Individual Multivariate Observations for Stability):
We assume that $X_1, X_2, \cdots, X_n$ are independently distributed as $N_p(\mu, \Sigma)$.
Then

$$X_j - \bar{X} = \left(1 - \frac{1}{n}\right)X_j - \frac{1}{n}X_1 - \frac{1}{n}X_2 - \cdots - \frac{1}{n}X_j - 1 - \frac{1}{n}X_{j+1} - \cdots - \frac{1}{n}X_n$$

Has

$$E\left(X_j - \bar{X}\right) = 0$$

And $\quad \mathrm{Cov}\left(X_j - \bar{X}\right) = \frac{(n-1)}{n}\Sigma$

Each $X_j - \bar{X}$ has a normal distribution, but, $X_j - \bar{X}$ is not independent of the sample covariance matrix S. However to set Control limits, we approximate that

$\left(X_j - \bar{X}\right)' S^{-1}\left(X_j - \bar{X}\right)$ has a chi-square distribution.

For the j^{th} point, we calculate the T^2-statistic

$$T_j^2 = \left(X_j - \bar{X}\right)' S^{-1}\left(X_j - \bar{X}\right)$$

We then plot the T^2-value on a time axis. The lower control limit is zero and we use the upper control limit

$UCL = \chi_p^{\,2}(0.05)$ at 95% level of significance or sometimes $\chi_p^{\,2}(0.01)$ for 99% level of significance.

There is no centerline in the T^2-chart.

Example 4 (A T^2-chart for Overtime Hours):

Using the following data, construct a T^2-chart on the two variables $X_1 =$ legal appearances hours and $X_2 =$ extraordinary event hours we take $\alpha = 0.01$ level of significance.

Table 2

No. of periods	X_1 (hrs.)	X_2 (hrs.)
1	3387	2200
2	3109	875
3	2670	957
4	3125	1758
5	3469	868
6	3120	398
7	3671	1603
8	4531	523
9	3678	2034
10	3238	1136
11	3135	5326
12	5217	1658
13	3728	1945
14	3506	344
15	3824	807
16	3516	1223

Solution: The T^2-chart for the data given in Table 2 is presented in Fig. 1, which reveals that the pair (legal appearances, extraordinary event) hours for period 11 is out of control. This is due to the large value of extraordinary event overtime during that period.

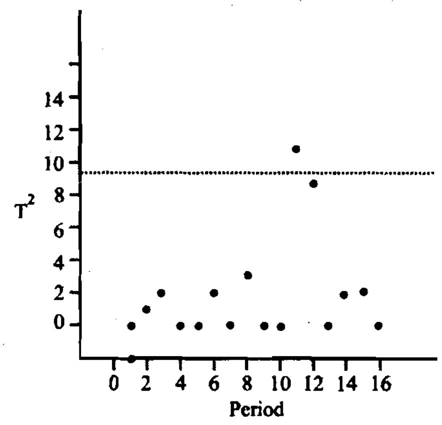


Fig. 1: T^2-Chart for legal appearance(X_1) and extraordinary event hour (X_2)

Now try the following exercises:

E4) Let $X_1, X_2, \cdots, X_{75}$ be a random sample from a population distribution with mean μ and covariance matrix Σ. What is the approximate distribution of each of the following?

a) $\overline{X}$ b) $n\left(\bar{X}-\mu\right)' S^{-1}\left(\bar{X}-\mu\right)$

E5) If X is distributed as $N_p(\mu, \Sigma)$ with $|\Sigma| \neq 0$, show that the joint density can be written as the product of marginal densities for $\underset{(q\times 1)}{X_1}$ and $\underset{((p-q)\times 1)}{X_2}$ if $\Sigma_{12} = \underset{(q\times(p-q))}{0}$

E6) Write true or false against the following statements. Give reasons also.

i) The maximum likelihood estimator of $\mu'\Sigma^{-1}\mu$ is $\hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}$, where $\hat{\mu} = \bar{X}$ and $\hat{\Sigma} = \left((n-1)/n\right)S$ are the maximum likelihood estimators of $\mu \,\&\, \Sigma$ respectively.

ii) Let $X_1, X_2, \cdots, X_n$ be a random sample from a multivariate normal population with mean μ and covariance Σ. Then $\bar{x}$ and S are sufficient statistics.

iii) Let $X_1, X_2, \cdots, X_n$ be independent observations from a population and mean μ and finite (nonsingular) covariance Σ. Then

a) $\sqrt{n}\left(\bar{X}-\mu\right)$ is approximately $N_p(0, \Sigma)$.

b) $n\left(\bar{X}-\mu\right)' S^{-1}\left(\bar{X}-\mu\right)$ is approximately $\chi_p^{\,2}$.

iv) Rejecting $H_0 : \mu = \mu_0$ when $|t|$ is large is equivalent to rejecting H_0 if it's square, $t^2 = \dfrac{\left(\bar{X}-\mu_0\right)^2}{S^2/n} = n\left(\bar{X}-\mu_0\right)\left(S^2\right)^{-1}\left(\bar{X}-\mu_0\right)$ is small.

v) If the observed statistical distance, T^2 is too large – that is if $\bar{X}$ is "too far" from μ_0 - the hypothesis $H_0 : \mu = \mu_0$ is accepted.

Let us now summarize the unit.

20.5 SUMMARY

In this unit, we have covered the following points.

1) Hotelling's T^2 is the test to which is applied to assess the statistical significance of the difference on the means of two or more variables between two groups.

2) T-test is used to test and assess the statistical significance of the difference between two independent sample means.

3) The Wishart distribution is a generalization of the gamma distribution to multiple dimensions. These distributions are of great importance in the estimation of covariance matrices in multivariate statistics.

4) A sampling distribution is the probability distribution, under repeated sampling of the population, of a given statistic. The formula for the sampling distribution depends on the distribution of the population, the statistic being considered, and the sample size used.

20.6 SOLUTIONS/ANSWERS

E1) a) We first find $\bar{X} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \end{bmatrix} = \begin{bmatrix} \frac{2+8+6+8}{2} \\ \frac{12+9+9+10}{2} \end{bmatrix} \bar{\bar{X}} = \begin{bmatrix} 12 \\ 20 \end{bmatrix}$

now,

$$S_{11} = \frac{(2-12)^2 + (8-12)^2 + (6-12)^2 + (8-20)^2}{3} = \frac{296}{3}$$

$$S_{22} = \frac{(12-20)^2 + (9-20)^2 + (9-20)^2 + (10-20)^2}{3} = \frac{406}{3}$$

$$S_{12} = \frac{(2-12)\,(12-20) + (8-12)\,(9-20) + (6-12)\,(9-2) + (8-20)(10-20)}{3}$$

$$S_{12} = \frac{310}{3}$$

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} \frac{296}{3} & \frac{310}{3} \\ \frac{310}{3} & \frac{406}{3} \end{bmatrix}$$

Therefore,

$$S^{-1} = \frac{1}{2675.11} \begin{bmatrix} \frac{406}{3} & -\frac{310}{3} \\ -\frac{310}{3} & \frac{296}{3} \end{bmatrix}$$

Since $T^2 = n\,(\bar{X} - \mu)^1\, S^{-1}\,(\bar{X} - \mu)$

For $n = 4$,

$$T^2 = 4[12-7,\, 20-11]\, S^{-1} \begin{bmatrix} 12-7 \\ 20-11 \end{bmatrix}$$

$$T^2 = 4[5,9]\, S^{-1} \begin{bmatrix} 5 \\ 9 \end{bmatrix} = 3.077$$

b) T^2 has the distribution of $\frac{(n-1)\,p}{n-p} F_{p,\,n-p}$

That is for $n = 4,\ p = 2$ (bivariate population), T^2 has

$\frac{(4-1).2}{(4-2)} F_{2,\,4-2} = 3F_{2,\,2}$ distribution.

E2) Calculate

$$T^2 = n(\bar{X} - \mu)^1\, S^{-1}\,(\bar{X} - \mu)$$

$$= 42[0.564 - 0.55,\, 0.603 - 0.60]\, S^{-1} \begin{bmatrix} 0.564 - 0.55 \\ 0.603 - 0.60 \end{bmatrix}$$

$T^2 = 1.169$ $\quad$ [S^{-1} is given]

Now T^2 has the distribution $\frac{(n-1)\,p}{(n-p)} F_{p,\,n-p}\,(\alpha)$

where $n = 42, p = 2, \alpha = 0.05$

Thus, $\frac{(42-1)\,2}{(42-2)} F_{2,\,40}\,(0.05)$

$= (2.05)\,(3.23)$

critical Value $= 6.62$

i.e. the critical value of T^2 is 6.62, which is more than the observed value of $T^2(1.169)$, so consequently, we accept H_0 at the 5% level of significance.

E3) Here $C = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, then

$CX_j = \begin{bmatrix} X_{ji} - X_{j2} \\ X_{j1} + X_{j2} \end{bmatrix}$ gives the new matrix,

$$(CX_j)' = \begin{bmatrix} (2-12) & (8-9) & (6-9) & (8-10) \\ (2+12) & (8+9) & (6+9) & (8+10) \end{bmatrix}'$$

$$\text{i.e., } CX_j \begin{bmatrix} (2-12) & (2+12) \\ (8-9) & (8+9) \\ (6-9) & (6+9) \\ (8-10) & (8+10) \end{bmatrix} = \begin{bmatrix} -10 & 14 \\ -1 & 17 \\ -3 & 15 \\ -2 & 18 \end{bmatrix}$$

using this, we can calculate $C\overline{X}$ and S^{-1} as has been computed in exercise 1(a). Accordingly,

$$T^2 = n(C\overline{X} - \mu)'\, S^{-1}\, (C\overline{X} - \mu).$$

The distribution of T^2 will remain unchanged that is $\frac{(n-1)\,p}{n-p} F_{p,\,n-p}(\alpha)$ for $n = 4$ and $p = 2$.

E4) a) Approximate distribution of $\overline{X}$ is $N_p\left(\mu, \left(\frac{1}{n}\right)\Sigma\right)$, where $n = 75$ (sample size).

b) Distribution of $n(\overline{X} - \mu)'\, S^{-1}\, (\overline{X} - \mu)$ is approximately χ_p^2 (chi-space) for $n-p$ large $(n-75)$.

E5) We can write,

$$(X-\mu)'\Sigma^{-1}(X-\mu) = [(X_1-\mu_1)(X_2-\mu_2)'] \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{bmatrix}$$

$$= (X_1-\mu_1)'\Sigma_{11}^{-1}(X_1-\mu_1) + (X_2-\mu_2)'\Sigma_{22}^{-1}(X_2-\mu_2)$$

where, $\Sigma = \begin{bmatrix} \Sigma_{11} & 0 \\ 0' & \Sigma_{22} \end{bmatrix}$ (Since $\Sigma_{12} = 0$)

$$\Sigma^{-1} = \frac{1}{\Sigma_{11}\Sigma_{22}} \begin{bmatrix} \Sigma_{22} & 0 \\ 0' & \Sigma_{11} \end{bmatrix} = \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0' & \Sigma_{22}^{-1} \end{bmatrix}$$

i.e., the joint density is the product of marginal densities for X_1 and X_2.

E6) i) True

ii) True

(Since it has been proved that the unbiased estimator of μ (for a normal population) is $\overline{X}$ and similarly, unbiased estimator of Σ is S. Moreover, all the information about μ and Σ is contained in $\overline{X}$ and S, regardless of the sample size n. Although, this is not true for non-normal populations. Thus, we can say that for normal populations, $\overline{X}$ and S are sufficient statistics for μ and Σ.)

iii) a) True

Since, we know that $\overline{X} \sim N_p(\mu, (1/n)\Sigma)$

Therefore, $\overline{X} \sim \mu N_p(0, (1/n)\,\Sigma)$

$$\Rightarrow \frac{\overline{X}-\mu}{\sqrt{\Sigma}/\sqrt{n}} \sim N_p(0, I_p)$$

or, $\sqrt{n}\,(\overline{X}-\mu) \sim N_p(0, \Sigma)$

b) True,

We start with,

$$n(\overline{X}-\mu)'\Sigma^{-1}(\overline{X}-\mu) = n(\overline{X}-\mu)'\Sigma^{-1/2}\Sigma^{-1/2}(\overline{X}-\mu)$$

Since $\overline{X} \sim N_p(\mu, (1/n)\,\Sigma)$

We can write $\frac{(\overline{X}-\mu)}{\sqrt{\Sigma}/\sqrt{n}} \sim N_p(0, I_p)$

where $Z = \sqrt{n}\,\Sigma^{-1/2}(\overline{X}-\mu)$ has a $N_p(0, I_p)$ distribution.

Now,

$$\begin{aligned} n(\overline{X}-\mu)'\Sigma^{-1}(\overline{X}-\mu) &= \sqrt{n}(\overline{X}-\mu)'\Sigma^{-1/2}\Sigma^{-1/2}(\overline{X}-\mu)\sqrt{n} \\ &= Z'Z \\ &= Z_1^2 + Z_2^2 + \cdots + Z_p^2 \end{aligned}$$

We now that χ_p^2 (chi-square) is defined as the distribution of the sum $Z_1^2 + Z_2^2 + \cdots + Z_p^2$, where $Z_1, Z_2, \ldots, Zp$ are independent $N(0, 1)$ random variables.

Therefore, $n(\overline{X}-\mu)\Sigma^{-1}(\overline{X}-\mu)$ is χ_p^2 distributed. Further, when n is large, S is close to Σ with maximum probability. Consequently, replacing Σ by S in $n(\overline{X}-\mu)'\,\Sigma^{-1}(\overline{X}-\mu)$ yields the outcome that $n(\overline{X}-\mu)'S^{-1}(\overline{X}-\mu)$ or approximately χ_p^2 distributed for n large.

iv) False

The test statistic $t = \frac{\overline{X}-\mu_0}{S/\sqrt{n}}$ for $H_0 : \mu = \mu_0$. This test statistic has a student's t distribution with $(n-1)$ d.f. We reject H_0 (null hypothesis) if the observed $|\,t\,|$ exceeds a specified value (critical value) of a t-distribution with $n-1$ degrees of freedom.

Now, $t^2 = \frac{(\overline{X}-\mu_0)^2}{S^2/n} = n(\overline{X}-\mu_0)\,(S^2)^{-1}(\overline{X}-\mu_0)$

Once $\overline{X}$ and S^2 are observed, we can state: Reject H_0 at significance level α if $n(\overline{X}-\mu_0)\,(S^2)^{-1}(\overline{X}-\mu_0) > t_{n-1}^2(\alpha/2)$ that is observed value of t^2 is more than the critical value of t^2, where $t_{n-1}(\alpha/2)$ denotes the upper $100(\alpha/2)^{th}$ percentile of the t-distribution. So in both the cases ((t) and t^2), the observed value and critical value of $|t|$ and t^2 respectively, are compared to reject or accept H_0 and since the fundamental principal remains the same in both the cases i we can reject H_0 if $|t|$ is large or its square, t^2 is large.

v) False

If the observed statistical distance T^2 is too large i.e., the calculated value of T^2 is too high than its critical value, then sample mean cannot be regarded as representative of population mean or we can say "$\overline{X}$ will be too far from μ_0". In that case the hypothesis $H_0 : \mu = \mu_0$ between sample mean and population mean is statistically significant.

20.7 PRACTICAL ASSIGNMENT

Session 5

1. Write a program in 'C' language to compute Hotelling T^2.