UNIT 22 INFERENCE-IV

Structure **Page No**

22.1 INTRODUCTION

Let $X(n \times p)$ be a data matrix observed for a random sample of size n from a multivariate normal population $N_p(\mu, \Sigma)$ where μ is the population mean vector and Σ is the population covariance matrix. In Unit 21, we have studied certain tests for testing the hypothesis regarding the population mean vector under the situation that covariance matrix is known as well as when it is unknown. We have also discussed tests for testing the equality of two mean vectors when two independent samples are available from multivariate normal populations. You may recall that the tests described there do depend on the assumption about the covariance matrices, e.g. whether the covariance matrices are known, unknown but equal, and unequal and unknown. Therefore, before performing the test for the mean vectors, we have to test certain hypotheses regarding the covariance matrices. The different cases of these will be discussed in Section 22.2, 22.3 and 22.4. In Section 22.5, we shall discuss the test for independence of components of multivariate random variables. In addition to the testing of hypotheses about the covariance matrices, another interest would be to test whether a number of independent samples available to us have come from same multivariate normal population or not i.e., we would like to test the homogeneity of covariance matrices which is done in Section 22.6 and the complete homogeneity of samples is done in Section 22.7. Tests for correlation matrices may be another interest which we shall discuss in Section 22.8. In this unit, we shall consider all these problems and corresponding tests would be developed.

Objectives

After reading this unit, you should be able to:

- formulate the null and alternation hypothesis for covariance matrices;
- derive a test statistics for testing the hypothesis regarding covariance matrices;
- use the tests for given data;
- test the homogeneity of covariance matrices;
- test for a correlation matrix.

22.2 TEST FOR EQUALITY OF ALL VARIANCES AND COVARIANCES

Let $X(n \times p)$ is a sample of size n drawn from a Normal population $N_p(\mu, \Sigma)$. Suppose we wish to test a null hypothesis that all variances and covariances are equal, i.e. matrix Σ is of a special form as given below:

$$\sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \cdots & \cdots & \cdots & \cdots \\ \rho & \rho & \cdots & 1 \end{bmatrix} = \sum_0 \quad \text{(say)} \tag{1}$$

The alternative hypothesis is that the covariance matrix $\Sigma \neq \Sigma_0$ is a general positive definite matrix.

We know that the likelihood function for the sample observation is

$$L(\mu, \Sigma|X) = \frac{\left|\Sigma^{-1}\right|^{n/2}}{(2\pi)^{np/2}} \exp\left[-\frac{1}{2}\sum_{\alpha=1}^{n}(X_\alpha - \mu)' \Sigma^{-1}(X_\alpha - \mu)\right]. \tag{2}$$

Let us denote by $X_{\alpha i}$, the i^{th} $(i = 1, 2, \ldots p)$ component of the α^{th} observation $(\alpha = 1, 2, \ldots n)$ in the sample, $X_\alpha = (X_{\alpha 1}, \cdots, X_{\alpha p})'$, $\mu = [\mu_1, \mu_2, \ldots \mu_p]'$ and $A = ((Aij)) = \Sigma^{-1}$. Thus, Eqn. (2) can be rewritten as

$$L(\mu, \Sigma|X) = \frac{|A|^{n/2}}{(2\pi)^{np/2}} \exp\left[-\frac{1}{2}\sum_{\alpha=1}^{n}\sum_{i,j=1}^{p} A_{ij}(X_{\alpha i} - \mu_i)(X_{\alpha j} - \mu_j)\right] \tag{3}$$

Under the null hypothesis $H_0 : \Sigma = \Sigma_0$, the matrix A should have a specific form having all its diagonal terms equal and similarly all its non-diagonal terms also equal to each other i.e.

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1p} \\ A_{21} & A_{22} & \cdots & A_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ A_{p1} & A_{p2} & \cdots & A_{pp} \end{bmatrix} = \begin{bmatrix} a & b & \cdots & b \\ b & a & \cdots & b \\ \cdots & \cdots & \cdots & \cdots \\ b & b & \cdots & a \end{bmatrix} \tag{4}$$

where $a = \dfrac{1 + (p-2)\rho}{\sigma^2(1-\rho)(1+(p-1)\rho}$ and $b = \dfrac{-\rho}{\sigma^2(1-\rho)(1+(p-1)\rho}$.

We know that the likelihood ratio test (LRT) is based on the ratio of maximum of likelihood function under the null hypothesis and over the whole space respectively i.e. the likelihood ratio for the present case is defined as

$$\lambda = \frac{\max\limits_{\mu, \Sigma_0} L(\mu, \Sigma|X)}{\max\limits_{\mu, \Sigma} L(\mu, \Sigma|X)} \tag{5}$$

The denominator of on right hand side of Eqn. (5) i.e. $\max\limits_{\mu, \Sigma} L(\mu, \Sigma|X)$, can be obtained by choosing the values of μ_i and A_{ij} such that it maximizes r.h.s. of Eqn. (3) and, hence, the values of μ_i and A_{ij} must satisfy the following $(p^2 + 3p)/2$ equations.

$$\frac{\partial L(\mu, \Sigma|X)}{\partial \mu_i} = 0 \quad i = 1, 2, \cdots, p \tag{6}$$

$$\frac{\partial L(\mu, \Sigma|X)}{\partial A_{ij}} = 0, i, j, = 1, 2, \ldots, p(i \le j). \quad (7)$$

These equations are respectively

$$n\left[\sum_{j=1}^{p} A_{ij}(\bar{X}_j - \mu_j)\right] L(\mu, \Sigma|X) = 0, \quad i = 1, 2, \ldots, p \quad (8)$$

$$\left[(n/2)A^{ij} - (1/2)\sum_{\alpha=1}^{n}(X_{\alpha i} - \mu_i)(X_{\alpha j} - \mu_j)\right]$$

$$L(\mu, \Sigma|X) = 0, i, j, = 1, 2, \ldots, p(i \le j) \quad (9)$$

where A^{ij} is the element in the i^{th} row and j^{th} column of $A^{-1} = \Sigma$, i.e., $A^{ij} = \rho i j \sigma_i \sigma_j$ and $\bar{X}_i = (1/n) \sum_{\alpha=1}^{n} X_{\alpha i}$. The solution of Eqn. (8) and Eqn. (9) is

$$\mu_i = \bar{X}_i \; i = 1, 2, \ldots, p \quad (10)$$

$$A^{ij} = s_{ij} \text{ or } A_{ij} = s^{ij} \; i, j, = 1, 2, \ldots, p(i \le j) \quad (11)$$

where $s_{ij} = (1/n)\sum_{\alpha-1}^{n}(X_{\alpha i} - \bar{X}_i)(X_{\alpha i} - \bar{X}_j), ((s_{ij})) = S$ (hence Eqn. (11) can also be written as $\Sigma = S$) and $((s_{ij}))^{-1} = ((s^{ij}))$. Substituting the values from Eqn.(10) and Eqn.(11) in Eqn.(3) and noting that the exponent in Eqn. (3) reduces to $\left(\frac{n}{2}\sum_{i,j=1}^{p} s^{ij}s_{ij}\right)$, which in turn reduces to $(-1/2)np$, since $\sum_{i=1}^{p} s^{ij}s_{ij} = 1$ for each value of j, we obtain

$$\max_{\mu,\Sigma} L(\mu, \Sigma|X) = \frac{\exp(-np/2)}{|S|^{n/2}(2\pi)^{np/2}}. \quad (12)$$

In order to obtain the numerator of right hand side of Eqn. (5) i.e. $\max_{\mu,\Sigma_0} L(\mu, \Sigma|X)$, we find μ_i's and A_{ij}'s so as to maximize the likelihood function in Eqn. (3) subject to the condition that $\Sigma = \Sigma_0$ i.e. the matrix A has the form given in Eqn. (4). Noting that the determinant A reduces to $\left[(a-b)^{p-1}(a + (p-1)b)\right]$, Eqn. (3) reduces to the following form

$$L^*(\mu, \Sigma|X) = \frac{[(a-b)^{p-1}(a + (p-1)b)]^{n/2}}{(2\pi)^{np/2}}$$

$$\exp\left[-\frac{1}{2}\left\{a\sum_{\alpha=1}^{n}\sum_{i=1}^{p}(X_{\alpha i} - \mu_i)^2 + b\sum_{\alpha=1}^{n}\sum_{i \ne j=1}^{p}(X_{\alpha i} - \mu_i)(X_{\alpha j} - \mu_j)\right\}\right] \quad (13)$$

The values of μ_i's, a and b which maximize L* are obtained by solving the following $(p+2)$ equations;

$$\frac{\partial L^*(\mu, \Sigma|X)}{\partial \mu_i} = 0, \quad i = 1, 2, \ldots, p;$$

$$\frac{\partial L^*(\mu, \Sigma|X)}{\partial a} = 0$$

and

$$\frac{\partial L^*(\mu, \Sigma|X)}{\partial b} = 0 \tag{14}$$

These equations are respectively

$$\left[(a-b)\sum_{\alpha=1}^{n}(X_{\alpha i}-\mu_i)+b\sum_{\alpha=1}^{n}\sum_{j=1}^{p}(X_{\alpha j}-\mu_j)\right]L^*(\mu,\Sigma|X)=0,\ i=1,2,\cdots,p \tag{15}$$

$$\left[\frac{n}{2}\left\{\frac{p-1}{a-b}+\frac{1}{a+(p-1)b}\right\}-\frac{1}{2}\sum_{\alpha=1}^{n}\sum_{i=1}^{p}(X_{\alpha i}-\mu_i)^2\right]L^*(\mu,\Sigma|X)=0 \tag{16}$$

and

$$\left[-\frac{n}{2}\left\{\frac{p-1}{a-b}-\frac{p-1}{a+(p-1)b}\right\}-\frac{1}{2}\sum_{\alpha=1}^{n}\sum_{i\neq j=1}^{p}(X_{\alpha i}-\mu_i)(X_{\alpha j}-\mu_j)\right]L^*(\mu,\Sigma|X)=0. \tag{17}$$

The above equations can further be simplified and written as follows:

$$\left[(a-b)(\overline{X}_i-\mu_i)+b\sum_{j=1}^{p}(\overline{X}_j-\mu_j)\right]=0,\ i=1,2,\ldots,p \tag{18}$$

$$\left[\left\{\frac{p-1}{a-b}+\frac{1}{a+(p-1)b}\right\}-\left\{ps^2+\sum_{i=1}^{p}(\overline{X}_i-\mu_i)^2\right\}\right]=0 \tag{19}$$

and

$$\left[\left\{\frac{p-1}{a-b}-\frac{p-1}{a+(p-1)b}\right\}+rp(p-1)s^2+\left(\sum_{i\neq j=1}^{p}(\overline{X}_i-\mu_i)(\overline{X}_j-\mu_j)\right)\right]=0 \tag{20}$$

where $r=\dfrac{\sum_{i\neq j=1}^{p}s_{ij}}{(p-1)\sum_{i=1}^{p}s_{ii}}$ and $s^2=\dfrac{\sum_{i=1}^{p}s_{ii}}{p}$. Solving Eqn. (18), we get $\mu_i=\overline{X}_i$ and substituting $\overline{X}_i$ for μ_i in Eqn. (19) and Eqn. (20), these reduce to

$$\frac{p-1}{a-b}+\frac{1}{a+(p-1)b}=ps^2 \tag{21}$$

and

$$\frac{-1}{a-b}+\frac{1}{a+(p-1)b}=rps^2 \tag{22}$$

respectively. Solving these equations for a and b, we get

$$a=\frac{1+(p-2)r}{s^2(1-r)(1+(p-1)r)} \tag{23}$$

and

$$b=\frac{-r}{s^2(1-r)(1+(p-1)r)}. \tag{24}$$

Substituting the values of μ_i, a and b in Eqn. (13), we get

$$\max_{\mu,\Sigma_0} L(\mu, \Sigma|X) = \frac{\exp\left(-\frac{1}{2}np\right)}{\left[(s^2)^p (1-r)^{p-1}(1+(p-1)r)\right]^{n/2} (2\pi)^{np/2}}. \quad (25)$$

Thus, substituting the values from Eqn. (12) and Eqn. (25) in Eqn. (5), we get the likelihood ratio for testing the hypothesis $H_0 : \Sigma = \Sigma_0$ as

$$\lambda = \left[\frac{|S|}{(s^2)^p (1-r)^{p-1}(1+(p-1)r)}\right]^{n/2} \quad (26)$$

We know that under $H_0, -2\log\lambda$ is approximately distributed as chi-square for large n. Therefore, Wilks (1946) suggested that if

$$\frac{|S|}{(s^2)^p (1-r)^{p-1}(1+(p-1)r)} = T, \quad (27)$$

$-n\log T$ would follow approximately chi-square distribution with $\{p(p+1)/2-2\}$ degrees of freedom under the null hypothesis. However Box (1949,1950) has shown that the following test statistic follows chi-square distribution with $\{p(p+1)/2-2\}$ degrees of freedom for large n:

$$\chi^2 = -\left[(n-1) - \frac{p(p+1)^2(2p-3)}{6(p-1)(p^2+p-4)}\right]\log\left(\frac{nT}{n-1}\right) \quad (28)$$

Therefore , we reject the null hypothesis if the calculated value of the test statistics as given in Eqn. (28) is greater than the tabulated value of chi-square with $\{p(p+1)/2-2\}$ degrees of freedom and at specified level of significance.

Now, let us illustrate it in the following example.

Example 1: The body dimensions of a certain species have been recorded. The information on body length (L) and body weight (W) are given in Table 1:

Table 1

Body length L (in mm)	Body weight W (in gm)
41	3.3
45	2.1
45	2.6
50	2.5
50	2.9
53	5.8
56	7.1
35	2.5
40	2.5
61	8.7

Test the hypothesis that all variances are equal and all co-variances are equal in the variance-covariance matrix for the given data.

Solution: Here, $n = 10$ and $p = 2$

$$S = \begin{bmatrix} 56.422 & 13.398 \\ 13.398 & 4.838 \end{bmatrix}$$

$$|S| = 93.463$$

Now, $T = \frac{|S|}{(s^2)^p (1-r)^{p-1}(1+(p-1)r)}$

where $r = \frac{\sum_{i \neq j = 1}^{p} s_{ij}}{(p-1)\sum_{i=1}^{p} s_{ii}}$ and $s^2 = \frac{\sum_{i=1}^{p} s_{ii}}{p}$

Therefore, $s^2 = 30.630$ and $r = 0.437$

$T = 0.1231$ and $nT/(n-1) = .1367$

Substituting above in Eqn.(28) which is again given below

$$\chi^2 = -\left[(n-1) - \frac{p(p+1)^2(2p-3)}{6(p-1)(p^2+p-4)}\right]\log\left(\frac{nT}{n-1}\right)$$

We get, $\chi^2 = -\left[9 - \frac{2(2+1)^2(2x2-3)}{6(2-1)(2^2+2-4)}\right]\log(0.1367) = -\frac{15}{2}\log(0.1367) = 6.481$.

Since the tabulated value of chi-square with $[p(p+1)/2-2]$ i.e. one degree of freedom at 5% level of significance is 3.84 which is less than the calculated value of the test statistics, we may reject the null hypothesis. Therefore, we may conclude that the body length and body weight for the given data are not having the same variances and same covariances.

Remark: One would like to test the hypothesis that all the components of mean vector are also equal along with the equality of variances and covariances i.e.

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_p (= \mu \text{ say}), \sigma_1 = \sigma_2 = \cdots = \sigma_p \text{ and } \rho_{12} = \rho_{13} = \cdots = \rho_{p-1,p}$$

The likelihood ratio for testing the above hypothesis is defined as follows

$$\lambda^* = \frac{\max_{H_0} L(\mu, \Sigma|X)}{\max_{\mu,\Sigma} L(\mu, \Sigma|X)} \quad (29)$$

The denominator of the above is same as obtained above in Eqn.(12) i.e.

$$\max_{\mu,\Sigma} L(\mu, \Sigma|X) = \frac{\exp(-np/2)}{|S|^{n/2}(2\pi)^{np/2}}. \quad (30)$$

Under the null hypothesis H_0 the likelihood takes the following form

$$L(\mu, \Sigma|X) = \frac{[(a-b)^{p-1}(a+(p-1)b)]^{n/2}}{(2\pi)^{np/2}}$$

$$\exp\left[-\frac{1}{2}\left\{a\sum_{\alpha=1}^{n}\sum_{i=1}^{p}(X_{\alpha i}-\mu)^2 + b\sum_{\alpha=1}^{n}\sum_{i \neq j=1}^{p}(X_{\alpha i}-\mu)(X_{\alpha j}-\mu)\right\}\right] \quad (31)$$

where a and b has been defined in Eqn.(4). The value of μ, a and b, which maximizes the above likelihood, is obtained by solving the following three equations:

$$\frac{\partial L(\mu, \Sigma|X)}{\partial \mu} = 0, \frac{\partial L(\mu, \Sigma|X)}{\partial a} = 0 \text{ and } \frac{\partial L(\mu, \Sigma|X)}{\partial b} = 0, \quad (32)$$

These equations simplify as follows

$$\left[(a-b)\sum_{\alpha=1}^{n}\sum_{i=1}^{p}(X_{\alpha i}-\mu)+b\sum_{\alpha=1}^{n}\sum_{i=1}^{p}(X_{\alpha i}-\mu)\right]L(\mu,\Sigma|X)=0,\ i=1,2,\ldots,p \qquad (33)$$

$$\left[\frac{n}{2}\left\{\frac{p-1}{a-b}+\frac{1}{a+(p-1)b}\right\}-\frac{1}{2}\sum_{\alpha=1}^{n}\sum_{i=1}^{p}(X_{\alpha i}-\mu)^2\right]L(\mu,\Sigma|X)=0 \qquad (34)$$

and

$$\left[-\frac{n}{2}\left\{\frac{p-1}{a-b}-\frac{p-1}{a+(p-1)b}\right\}-\frac{1}{2}\sum_{\alpha=1}^{n}\sum_{i\neq j=1}^{p}(X_{\alpha i}-\mu)(X_{\alpha i}-\mu)\right]L(\mu,\Sigma|X)=0 \qquad (35)$$

Solving these equations, we get

$$\mu=\overline{X} \qquad (36)$$

$$a=\frac{1+(p-2)r_0}{s_0^2(1-r_0)(1+(p-1)r_0)} \qquad (37)$$

and

$$b=\frac{-r_0}{s_0^2(1-r_0)(1+(p-1)r_0)}. \qquad (38)$$

where,

$$\overline{X}=\frac{1}{np}\sum_{\alpha=1}^{n}\sum_{i=1}^{p}X_{\alpha i}$$

$$s_{0ij}=\frac{1}{n}\sum_{\alpha=1}^{n}(X_{\alpha i}-\overline{X})(X_{\alpha j}-\overline{X})=s_{ij}+(\overline{X}_i-\overline{X})(\overline{X}_j-\overline{X})$$

$$r_0=\frac{\sum_{i\neq j=1}^{p}s_{0ij}}{(p-1)\sum_{i=1}^{p}s_{0ii}}=\frac{\sum_{i\neq j=1}^{p}s_{ij}-\sum_{i=1}^{p}(\overline{X}_i-\overline{X})}{(p-1)\sum_{i=1}^{p}s_{ii}+\sum_{i=1}^{p}(\overline{X}_i-\overline{X})^2}$$

and

$$s_0^2=\frac{1}{p}\sum_{i=1}^{p}s_{0ii}.$$

Thus, we get

$$\max_{H_0}L(\mu,\Sigma|X)=\frac{\exp\left(-\frac{1}{2}np\right)}{\left[(s_0^2)^p(1-r_0)^{p-1}(1+(p-1)r_0)\right]^{n/2}(2\pi)^{np/2}} \qquad (39)$$

Substituting the values from Eqn. (30) and Eqn. (39) in Eqn. (29) we get the likelihood ratio for testing the hypothesis H_0 as

$$\lambda^*=\left[\frac{|S|}{(s_0^2)^p(1-r_0)^{p-1}(1+(p-1)r_0)}\right]^{n/2} \qquad (40)$$

and $-n\log T^*$, where

$$T^*=\frac{|S|}{(s_0^2)^p(1-r_0)^{p-1}(1+(p-1)r_0)}, \qquad (41)$$

is distributed as a chi-square with $\left(\frac{p(p+3)}{2}-3\right)$ degrees of freedom under the null hypothesis. Therefore null hypothesis is rejected if $-n \log T^*$ greater than tabulated x_2 value with $\left\{\frac{p(p+3)}{2}-3\right\}$ degrees of freedom.

Example 2: A group of 100 students appeared in three consecutive tests and the scores obtained by them are noted down to check whether the tests can be considered as similar. If the tests are similar, it is expected that the students' score in three tests should have same means, variances and covariances. Thus, the test discussed in the above remark can be applied. From the data, the various statistics needed for the test are calculated and these are given below:

$$\overline{X_1}=10.9900\ \overline{X_2}=10.9300\ \overline{X_3}=11.2600$$

$$S=\begin{bmatrix}16.8451 & 13.5493 & 14.5826\\ 13.5493 & 18.1099 & 13.8056\\ 14.5826 & 13.8056 & 17.7124\end{bmatrix} \text{and} |S|=545.5308$$

$$s_0^2=17.5764, |S|=545.5308, r_0=0.7984$$

Since, $n=100$ and $p=3$, it may be verified that the value of the test statistic T^* given in Eqn. (41) comes out to be 0.9370. Further $-n\log T^*=6.5072$ and the tabulated value of chi-square with 6 degrees of freedom at 5% level of significance is 12.59. The calculated value of the statistics is less than the tabulated value, therefore we may conclude that the data do not provide enough evidence for the rejection of null hypothesis and hence we may accept that the scores of the students in the three tests have same means, variances and covariances.

Now try the following exercise.

E1) The body dimensions of 4 different species of insects collected during February-March, 1993 have been recorded [Bhuyan and Nair (1995)]. The information on body length (L) and body weight (W) of 20 insects of each species are given below:

Specie I		Specie II		Specie III		Specie IV	
L(mm)	W(gm)	L(mm)	W(gm)	L(mm)	W(gm)	L(mm)	W(gm)
36	1.3	41	3.3	45	1.8	27	0.8
37	1.8	45	2.1	45	1.7	28	0.5
37	1.7	45	2.6	46	1.9	35	1.6
40	2.4	45	3.8	49	2.0	30	0.5
41	2.5	45	6.9	49	2.3	45	2.8
42	2.3	45	4.4	50	2.2	42	1.3
43	2.5	50	2.5	50	2.4	45	2.4
44	2.2	50	2.9	51	2.0	40	1.3
46	2.7	50	2.6	52	2.1	45	2.8
47	3.3	50	3.1	53	2.1	48	3.4
48	2.2	52	5.3	54	3.2	45	2.9
48	2.9	53	5.8	53	2.1	48	2.4
51	3.1	54	3.3	55	3.0	45	2.8
52	3.3	55	5.9	55	2.2	48	2.9
53	3.6	56	7.1	55	2.5	44	2.4
59	2.9	57	8.0	56	3.4	45	2.3
55	4.5	35	2.5	57	3.3	45	3.1
55	3.6	40	1.7	57	3.4	42	1.7
56	5.0	40	2.5	58	3.1	50	2.4
60	4.5	61	8.7	50	2.5	52	3.7

Use the above data to test the hypothesis that variances and covariances for three species are equal.

In the following section we shall discuss the test for covariance matrix to be equal to a given matrix.

22.3 TEST FOR COVARIANCE MATRIX TO BE EQUAL TO A GIVEN MATRIX

Let $X(nxp)$ is a sample of size n drawn from a Normal population $N_p(\mu, \Sigma)$. Suppose we wish to test the null hypothesis that covariance matrix is equal to the given matrix i.e. $H_0 : \Sigma \neq \Sigma_0$, against the alternative hypothesis $H_A : \Sigma \neq \Sigma_0$. It is also assumed that vector μ is unknown.

The likelihood ratio for testing the above hypothesis is as follows

$$\lambda = \frac{\max\limits_{\mu,\Sigma_0} L(\mu, \Sigma|X)}{\max\limits_{\mu,\Sigma} L(\mu, \Sigma|X)} \qquad (42)$$

The likelihood function for the sample observation is given in Eqn. (2) which can be rewritten as

$$L(\mu, \Sigma | X) = \frac{|\Sigma^{-1}|^{n/2}}{(2\pi)^{np/2}} \left[\exp -\frac{1}{2} \operatorname{tr} \Sigma^{-1} \left\{ \sum_{\alpha=1}^{n} (X_\alpha - \mu)(X_\alpha - \mu)' \right\} \right] \qquad (43)$$

Note: The trace of a (pxp) square matrix $A = ((Aij))$ is defined as tr $A = \sum_{i=1}^{p} A_{ii}$. The power of the exponential term in Eqn. (2) is 1×1, hence it will be equal to its trace. The following are the properties of trace:

$$\operatorname{tr}(A + B) = \operatorname{tr} A + \operatorname{tr} B$$

$$\operatorname{tr} AB = \operatorname{tr} BA$$

We know that the maximum likelihood estimator of μ is $\overline{X}$ and that of Σ is S, where S is the sample variance-covariance matrix. Hence it is easy to verify that

$$\max_{\mu,\Sigma_0} L(\mu, \Sigma | X) = \frac{|\Sigma_0^{-1}|^{n/2}}{(2\pi)^{np/2}} \exp\left[-\frac{n}{2} \operatorname{tr} \Sigma_0^{-1} S \right] \qquad (44)$$

and

$$\max_{\mu,\Sigma} L(\mu, \Sigma | X) = \frac{|S^{-1}|^{n/2}}{(2\pi)^{np/2}} \exp\left[-\frac{np}{2} \right] \qquad (45)$$

substituting Eqn. (44) and Eqn. (45) in Eqn. (42), we have

$$\lambda = \left|\Sigma_0^{-1} S\right|^{n/2} \exp\left[-\frac{n}{2} \operatorname{tr}\Sigma_0^{-1} S + \frac{np}{2} \right]$$

and hence

$$-2\log\lambda = -n \log\left|\Sigma_0^{-1} S\right| + n \operatorname{tr} \Sigma_0^{-1} S - np \qquad (46)$$

This statistics is a function of the eigen values of $\Sigma_0^{-1}S$. Let a and g be the arithmetic and geometric mean of eigen values of $\Sigma_0^{-1}S$. Then tr $\Sigma_0^{-1}S = pa$ and $\left|\Sigma_0^{-1}S\right| = g^p$. Therefore Eqn. (47) can also be written as

$$-2\log\lambda = np(a - \log(g) - 1) \qquad (47)$$

It may be recalled that the asymptotic distribution of $-2\log\lambda$ is chi-square distribution with $p(p+1)/2$ degrees of freedom and hence H_0 is rejected when $-2\log\lambda$ given by Eqn.(46) (or Eqn.(47)) is greater than tabulated χ^2 value with $\frac{p(p+1)}{2}$ degrees of freedom.

Example 3: The data given in Example 1 is a part of information. For the complete data set the population variance-covariance matrix may be as given below:

$$\Sigma_0 = \begin{bmatrix} 20.421 & 2.582 \\ 2.582 & 1.838 \end{bmatrix}.$$

Let us investigate whether or not the data given in the Example 1 confirms to the hypothesis that it has been drawn from a bivariate normal population having variance covariance matrix as Σ_0, i.e. we wish to test the hypothesis $H_0 : \Sigma = \Sigma_0$ against the alternative hypothesis $H_1 : \Sigma \neq \Sigma_0$. From the Example 1 you may note that

$$S = \begin{bmatrix} 56.422 & 13.398 \\ 13.398 & 4.838 \end{bmatrix}.$$

Hence,

$$\Sigma_0^{-1}S = \begin{bmatrix} 0.0595 & -0.0836 \\ -0.0836 & 0.6615 \end{bmatrix}\begin{bmatrix} 56.422 & 13.398 \\ 13.398 & 4.838 \end{bmatrix} = \begin{bmatrix} 2.237 & 0.3927 \\ 4.0947 & 2.0972 \end{bmatrix}.$$

The calculated value of the test statistic $-n\log + n\,\mathrm{tr}\,\Sigma_0^{-1}S - np\left|\Sigma_0^{-1}S\right|$ is, therefore obtained as $-10\log(3.0835) + 10(4.3352) - 10\times 2 = 22.8529$.

Since the calculated value of the test statistics which follows chi-square distribution under null hypothesis with $p(p+1)/2$ (3 in this case) degrees of freedom, is greater than the tabulated value (7.81) of chi-square with 3 degrees of freedom at 5% level of significance, we reject H_0. Hence we conclude that covariance matrix for the length and weight data differs from the assumed covariance matrix Σ_0.

Now try an exercise.

E 2) Let us consider the data given in E1) and test the hypothesis that

$$\Sigma = \begin{bmatrix} 19.63 & 1.46 \\ 1.46 & 4.31 \end{bmatrix}.$$

In the following section we shall discuss the test for the covariance matrix to be equal to constant multiple of a given matrix.

22.4 TEST FOR COVARIANCE MATRIX TO BE EQUAL TO CONSTANT MULTIPLE OF A GIVEN MATRIX

Let X $(n\times p)$ is a sample of size n drawn from a Normal population $N_p(\mu, \Sigma)$. Let us consider that we wish to test the null hypothesis that covariance is equal to constant multiple of a given matrix i.e. $H_0 : \Sigma = k\Sigma_0$ against the alternative hypothesis $H_A : \Sigma \neq k\Sigma_0$, where k is some unknown constant. It is also assumed that μ is unknown.

The likelihood ratio for testing the above hypothesis is as follows

$$\lambda = \frac{\max_{H_0} L(\mu, \Sigma|X)}{\max_{H_A} L(\mu, \Sigma|X)}. \quad (48)$$

The likelihood function for the sample observation is given in Eqn. (43). Therefore we have

$$\log L(\mu, \Sigma|X) = -\frac{np}{2}\log 2\pi - \frac{n}{2}\log|\Sigma| - \frac{n}{2}\mathrm{tr}\{\Sigma^{-1}S + (\overline{X}-\mu)'\Sigma^{-1}(\overline{X}-\mu)\} \quad (49)$$

Under H_0 it reduces to

$$\log L(\mu, k\Sigma_0|X) = -\frac{np}{2}\log 2\pi - \frac{n}{2}\log$$

$$|k\Sigma_0| - \frac{n}{2}k^{-1}\mathrm{tr}\{\Sigma_0^{-1}S + (\overline{X}-\mu)'\Sigma_0^{-1}(\overline{X}-\mu)\} \quad (50)$$

Here $\overline{X}$ is sample mean vector $\overline{X} = (\overline{X}_1, \overline{X}_2, \cdots, \overline{X}_p)'$.

In order to obtain the numerator of Eqn. (48), we choose the values of μ and k such that Eqn. (50) is maximized. Such a value of μ and k can be obtained by solving the following equations:

$$\frac{\partial}{\partial \mu}\log L(\mu, k\Sigma_0|X) = nk^{-1}\Sigma_0^{-1}(\overline{X}-\mu) = 0 \quad (51)$$

and

$$\frac{\partial}{\partial k}\log L(\mu, k\Sigma_0|X) = \left(-\frac{n}{2}\right)\left[\left(\frac{p}{k}\right) + \frac{1}{k^2}\mathrm{tr}\left\{\Sigma_0^{-1}S + (\overline{X}-\mu)'\Sigma_0^{-1}(\overline{X}-\mu)\right\}\right] = 0 \quad (52)$$

Solving Eqn. (51) and Eqn. (52) simultaneously we get $\mu = \overline{X}$ and $k = \mathrm{tr}\ \Sigma_0^{-1}S/p$. Hence

$$\max_{H_0} L(\mu, \Sigma|X) = \frac{1}{(2\pi)^{np/2}|\hat{k}\Sigma_0|^{n/2}}\exp\left[\left(-\frac{1}{2}np\right)\right] \quad (53)$$

And

$$\max_{H_A} L(\mu, \Sigma|X) = \frac{1}{(2\pi)^{np/2}|S|^{n/2}}\exp\left[\left(-\frac{1}{2}np\right)\right] \quad (54)$$

Substituting Eqn. (53) and Eqn. (54) in Eqn. (48), we have

$$\lambda = \frac{|S|^{n/2}}{|\hat{k}\Sigma_0|^{n/2}} = \frac{|\Sigma_0^{-1}S|^{n/2}}{(\hat{k})^{np/2}} \quad (55)$$

$$\text{Thus,} -2\log\lambda = np\log\left(\frac{\hat{k}}{|\Sigma_0^{-1}S|^{1/p}}\right) = np\log\left(\frac{\mathrm{tr}\Sigma_0^{-1}S/p}{|\Sigma_0^{-1}S|^{1/p}}\right) = np\log(a/g) \quad (56)$$

where a and g are the arithmetic mean and geometric mean of eigen values of $\Sigma_0^{-1}S$, respectively. Since $-2\log\lambda$ is asymptotically distributed as chi-square with $(p-1)(p+2)/2$ degrees of freedom, we reject null hypothesis H_0 if calculated value of the statistics from Eqn. (56) is greater than the corresponding tabulated value of chi-square with $(p-1)(p+2)/2$ degrees of freedom at specified level of significance.

Example 4: In Example 3 we have seen that the data given in Example 1 do not confirm the assumption that it has come from the population with variance-covariance matrix

$$\Sigma_0 = \begin{bmatrix} 20.421 & 2.582 \\ 2.582 & 1.838 \end{bmatrix}$$

Let us now try to test whether or not the population variance-covariance matrix is a constant multiple of Σ_0. In other words, we wish to test the null hypothesis $H_0 : \Sigma = k\,\Sigma_0$ against the alternative hypothesis $H_1 : \Sigma \neq k\,\Sigma_0$. The test described above can be used for this purpose. It may be noted from Example 1 that the sample variance-covariance matrix is

$$S = \begin{bmatrix} 56.422 & 13.398 \\ 13.398 & 4.838 \end{bmatrix}.$$

Further note from Example 3 that

$$\Sigma_0^{-1}S = \begin{bmatrix} 2.237 & 0.3927 \\ 4.0947 & 2.0972 \end{bmatrix}.$$

Here, $n = 10$ and $p = 3$ and hence, the estimate of k

$$\hat{k} = p^{-1}\mathrm{Tr}(\Sigma_0^{-1}S) = 2.1671$$

$$\text{And} = \left|\Sigma_0^{-1}S\right|^{1/p} = 1.7559$$

The value of the test statistic $np\log\ \hat{k}\big/\left(\Sigma_0^{-1}S\right)^{1/p}$ which distributed as chi square with $(p-1)(p+2)/2$ degrees of freedom is $20\log(1.2341) = 1.870$.

Since the calculated value of the test statistics is less than the tabulated value (5.99) of chi square with 2 degrees of freedom at 5% level of significance, hence we cannot reject the null hypothesis. Therefore we may conclude that the population covariance is constant multiple of matrix Σ_0.

Now try the exercise.

E 3) Let us consider E 1) and test the hypothesis that variance-covariance matrix is constant multiple of $\begin{bmatrix} 19.63 & 1.46 \\ 1.46 & 4.31 \end{bmatrix}$.

Let us now discuss the test for independence of components of variables in the following section.

22.5 TEST FOR INDEPENDENCE OF COMPONENTS OF VECTOR RANDOM VARIABLE

Let $X(n \times p)$ is a sample of size n drawn from a Normal population $N_p(\mu, \Sigma)$. Suppose that each observed vector is partitioned in two parts, one consisting of p_1 and the other p_2 components such that $p = p_1 + p_2$. The likelihood function of X can also be split into two factors, one for parameters μ_1 and Σ_{11}, and an other for μ_2 and Σ_{22}, where, $\mu = \left[\mu_1' \ \mu_2'\right]$ and $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$. Let us consider that we wish to test that the sets of variables corresponding to first p_1 components are independent to the set of variables corresponding to remaining p_2 components of the observed vector i.e. $H_0 : \Sigma_{12} = 0$, against the alternative hypothesis $H_A : \Sigma_{12} \neq 0$.

Under $H_0, \Sigma_{12} = 0$, and the MLE of Σ is $\hat{\Sigma} = \begin{bmatrix} S_{11} & 0 \\ 0 & S_{22} \end{bmatrix}$.

Also $\hat{\mu} = \begin{bmatrix} \hat{\mu}_1', \hat{\mu}_2' \end{bmatrix}' = \overline{X}$. Here MLE of μ under both H_0 and H_A is $\overline{X}$. Comparing with discussion in earlier sections we now see that the LRT for the hypothesis depends on the eigenvalues of $\hat{\Sigma}^{-1} S$, i.e. $\Sigma^{-1} S = \begin{bmatrix} S_{11} & 0 \\ 0 & S_{22} \end{bmatrix}^{-1} \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} I & S_{11}^{-1} S_{12} \\ S_{22}^{-1} S_{21} & I \end{bmatrix}$

Here tr $\Sigma^{-1} S = p$ and hence the arithmetic mean of eigenvalues is one. The geometric mean is obtained as

$$g^p = \left| \hat{\Sigma}^{-1} S \right| = |S| / (|S_{11}||S_{22}|) = \left| S_{22} - S_{21} S_{11}^{-1} S_{12} \right| / |S_{22}|$$

Now, according to the LRT criterion, the test statistics (similar to Eqn.(47)) is

$$-2 \ln \lambda = -n \ln \left(\left| S_{22} - S_{21} S_{11}^{-1} S_{12} \right| / |S_{22}| \right) = -n \ln \left| I - S_{22}^{-1} S_{21} S_{11}^{-1} S_{12} \right|$$

$$= -n \ln \prod_{i=1}^{k} (1 - \lambda_i) \qquad (57)$$

where $k = \min(p_1, p_2)$ and λ_i's are the non zero eigenvalues of $S_{22}^{-1} S_{21} S_{11}^{-1} S_{12}$. The test statistic can also be written in terms of correlation matrix i.e.

$$-2 \ln \lambda = -n \ln \left| I - R_{22}^{-1}, R_{21} R_{11}^{-1} R_{12} \right|.$$

Finally, the test statistic can be written as:

$$\lambda^{2/n} = \prod_{i=1}^{k} (I - \lambda_i).$$

Bartlett (1954) has shown that

$$-\left[n - \frac{1}{2}(p_1 + p_2 + 3) \right] \ln \lambda^{2/n} \sim \chi^2_{p_1 p_2}$$

for larger n such that $n - 1 \geq p_1 + p_2$.

If $p_1 = 1$ and $p_2 = p - 1$, the test statistic takes the form of

$-2 \ln \lambda = -n \ln (1 - R^2)$

where R is the multiple correlation coefficient between the first variable and the others.

Now try the following exercise.

E 4) The following information relate with fertility level and child mortality level along with some socio-economic variable of couples of child bearing ages in a country according to their socio-economic status:

Objects	**No. of ever born children** x_1	**No. of dead children under age 5** x_2	**Level of education (in Complete years)** Father x_3	Mother x_4	**Duration of marriage (in years)** x_5
1.	10	3	5	0	18
2.	3	0	10	5	8
3.	5	0	12	4	15

4.	2	0	14	14	6
5.	12	2	0	0	17
6.	1	0	10	0	3
7.	8	1	5	5	15
8.	7	2	0	0	18
9.	4	0	8	4	10
10.	2	0	12	10	5
11.	1	0	10	5	3
12.	5	1	5	0	12
13.	4	0	0	0	11
14.	6	2	5	0	13
15.	7	1	0	0	10
16.	4	1	6	0	12
17.	8	2	0	0	15
18.	3	1	0	0	8
19.	6	1	0	0	14
20.	5	0	6	0	12
21.	2	0	17	12	5
22.	5	0	12	12	12
23.	4	1	10	10	8
24.	6	1	8	0	14
25.	3	0	12	10	8
26.	4	0	10	5	17
27.	5	1	6	6	8
28.	10	2	6	6	19
29.	8	3	6	6	12
30.	6	1	10	6	14
31.	7	0	10	0	15
32.	6	0	16	10	14
33.	5	0	16	12	13
34.	4	0	10	10	12
35.	8	1	0	0	15
36.	7	2	0	0	13
37.	3	0	0	0	8
38.	2	0	16	16	5
39.	3	0	16	16	10
40.	1	0	17	17	4
41.	2	0	17	17	6
42.	4	1	10	10	9
43.	5	1	10	10	18
44.	4	0	10	10	12
45.	6	0	8	8	15
46.	4	0	10	10	12
47.	6	0	12	12	17
48.	5	0	10	10	15
49.	4	0	12	12	10
50.	3	0	10	10	12
51.	7	1	10	10	11
52.	8	4	6	6	12
53.	2	0	17	17	6
54.	1	0	17	12	3
55.	4	0	10	10	12
56.	5	0	12	12	16
57.	3	0	10	10	10
58.	4	0	6	6	16
59.	6	1	8	0	14

60.	5	0	6	0	15
61.	1	0	12	12	4
62.	2	0	12	10	3
63.	1	0	12	12	2
64.	2	0	10	10	5

Test for independence of number of ever born children (x_1) and number of dead children (x_2) with mother's education (x_4) and duration of marriage (x_5) using above data.

Now we shall discuss the test of homogeneity of covariance matrices.

22.6 TEST OF HOMOGENEITY OF COVARIANCE MATRICES

Suppose there are K multivariate normal populations $N_p(\mu_i, \Sigma_i)$ $i = 1, 2, \cdots, k$. Let $X_i(n_i \times p)$ be data matrix from $N_p(\mu_i, \Sigma_i)$ $[i = 1, 2, \ldots, k,]$ and we want to test the hypothesis that k covariance matrices are equal i.e. $H_0 : \Sigma_1 = \Sigma_2 = \ldots = \Sigma_k$ against the alternative hypothesis that all the co-variance matrices are not equal i.e. $H_A : \Sigma_i \neq \Sigma_j$ for at least one $i \neq j$, $i, j = 1, 2, \ldots, k$.

Under H_0 and H_A the MLE of μ_i is $\overline{X_i}$.

Under H_0 the MLE of common covariance matrix Σ is $\frac{w}{n}$ where $n = \sum_{i=1}^{k} n_i$ and $w = \sum n_i S_i$, and S_i is the MLE of Σ_i under H_A. Then the test statistic for the null hypothesis is

$$-2\ln\lambda = n\ln|S| - \sum n_i \ln|S_i|, \tag{58}$$

where $S = n^{-1}w$. This statistic has an asymptotic Chi-square distribution with $\frac{1}{2}p(p+1)(k-1)$ degrees of freedom. Box (1949) proposed M-statistic, where

$$M = (n-k)\ln|S_u| - \sum_{i=1}^{k}(n_i - 1)\ln|S_{iu}|,$$

$$S_u = \frac{nS}{n-k} \text{ and } S_{iu} = \frac{n_i S_i}{n_i - 1} \tag{59}$$

According to Box MC^{-1} has a Chi-Square distribution with $\frac{1}{2}p(p+1)(k-1)$ degrees of freedom, where

$$C^{-1} = 1 - \frac{2p^2 + 3p - 1}{6(p+1)(k-1)}\left[\sum_{i=1}^{k}\left\{\frac{1}{n_i - 1} - \frac{1}{n-k}\right\}\right]. \tag{60}$$

If all n_i are equal, then

$$C^{-1} = 1 - \frac{(2p^2 + 3p - 1)(k+1)}{6(p+1)(n-k)k} \tag{61}$$

The above MC^{-1} statistic is known as Box's M-test statistic. The approximation of MC^{-1} statistic is good if each n_i, exceeds 20, and p and k do not exceed 5. For p and k greater than $5, n_i < 20$, Box has suggested an asymptotic F distribution [Pearson and Hartley (1972)]. Roy, Pillai and others have proposed alternative test procedures for $k = 2$.

Example 5: To illustrate the above test procedure let us consider the data on the body dimension of three species X_1, X_2 and X_3. The body length and body weight for these are given below:

X_1		X_2		X_3	
Body length L (mm)	**Body weight W (gm)**	**Body length L (mm)**	**Body weight W (gm)**	**Body length L (mm)**	**Body weight W (gm)**
41	3.3	45	1.8	27	0.8
45	2.1	46	1.9	35	1.6
45	2.6	49	2.3	30	0.5
50	2.5	50	2.4	45	2.8
50	2.9	52	2.1	42	1.3
53	5.8	53	2.1	40	1.3
56	7.2	54	3.2	48	3.4
35	2.5	53	2.1	45	2.9
40	2.5	58	3.1	44	2.4
61	8.7	50	2.5	42	1.7

We wish to test the homogeneity of variance-covariance matrices for the three species i.e., our null hypothesis is $H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3$ against the alternative hypothesis H_A : At least two out of the three Σ_i 's differ. Here $k = 3, n_i = 10(i = 1,2,3)$ and $p = 2$. From the data it is easy to calculate

$$S_{1u} = \begin{bmatrix} 56.422 & 13.398 \\ 13.398 & 4.838 \end{bmatrix}$$

$$S_{2u} = \begin{bmatrix} 13.40 & 1.20 \\ 1.20 & 0.20 \end{bmatrix}$$

$$S_{3u} = \begin{bmatrix} 4.160 & 5.164 \\ 5.164 & 0.832 \end{bmatrix}$$

Hence, $|S_{1u}| = 93.463, |S_{2u}| = 1.24$ and $|S_{3u}| = 9.243$, $S_u = \begin{bmatrix} 112.982 & 20.762 \\ 20.762 & 5.270 \end{bmatrix}$ and hence,

$|S_u| = 164.3545$.

Substituting the above values in Eqn. (60), we get $M = 19.1535$ and from Eqn. (62), we get $C^{-1} = 0.8395$. The value of MC^{-1} is, therefore, obtained as 16.07944. The tabulated value of chi-square with 6 degrees of freedom at 5% level of significance is 12.59. Since, calculated value of the test statistics is greater than tabulated value, we may reject the null hypothesis at 5% level of significance and conclude that the covariance matrices are not homogeneous.

Try an exercise.

E5) Consider E1) and test the homogeneity of covariance matrixes for species I, II, III and IV between their length and weight.

In the following section, test of complete homogeneity is discussed.

22.7 TEST OF COMPLETE HOMOGENEITY

Let $X_i(n_i \times p)$ be data matrix from $N_p(\mu_i, \Sigma_i)[i = 1,2,\ldots,k]$. For such data matrices the LRT to test $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$ has been discussed in the previous unit i.e. Unit 21 and the test statistics is

$$\lambda = \left|I + W^{-1}B\right|^{-n/2}$$

under the assumption that $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_k$. Again, for the hypothesis $H_0 : \Sigma_1 = \Sigma_2 = \ldots = \Sigma_k$ the test statistic as discussed in Section 22.6 is

$$-2\ln\lambda_1 = (n-k)\ln|S_u| - \sum_{i=1}^{k}(n_i - 1)\ln|S_{iu}|$$

where $n = \sum_{i=1}^{k} n_i$. Now, the problem is to test the hypothesis $H_0 : \mu_1 = \ldots. = \mu_k$ and $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_k$. The LRT statistic for this latter hypothesis is $-2\ln\lambda_2 = -2\ln\lambda_1 - 2\ln\lambda$

This statistic is asymptotically distributed as χ^2 with $\frac{1}{2}p(k-1)(p+3)$ degrees of freedom. On simplification the statistic stands as

$$-2\ln\lambda_2 = (n-k)\ln|W_1| - \sum_{i=1}^{k}(n_i - 1)\ln|S_{iu}| \qquad (62)$$

where $S_{iu} = \frac{n_i S_i}{n_i - 1}$ is the unbiased estimator of Σ_i and $W_1 = \frac{\Sigma(n_i - 1)S_{iu}}{n-k}$.

After this we shall discuss the test of correlation matrix.

22.8 TEST OF CORRELATION MATRIX

Let $X(n \times p)$ be a data matrix from $N_p(\mu, \Sigma)$. Assume that the population correlation matrix of p variable is P. We need to test of hypothesis $H_0 : P = I$ against $H_A : P \neq I$ The null hypothesis indicates that all the p variable is uncorrelated. Under this hypothesis, the mean and variance of each variable are estimated separately. The ML estimates provide $\hat{\mu} = \bar{X}$ and $\hat{\Sigma} = \text{diag}(s_{11} \ldots. s_{pp})$, where s_{ii} is the sample variance of i^{th} variable. Let the sample correlation matrix be R.

The above hypothesis is equivalent to $H_0 : \Sigma = \text{diag}(\sigma_{11}\, \sigma_{22} \ldots \sigma_{pp})$, a special form of the Σ matrix. The LRT statistic for such a hypothesis is equivalent that of $H_0 : \Sigma = \Sigma_0$. Now, using R matrix instead of S matrix, the LRT statistic for the required hypothesis is

$$-2\ln\lambda = -n\ln|R| \qquad (63)$$

where $-2\ln\lambda$ is asymptotically distributed as Chi-square with $\frac{1}{2}p(p-1)$ degrees of freedom. Box (1949) has shown that if n is replaced by $n' = n - \frac{1}{6}(2p+11)$, then chi-square approximation is improved.

Example 6: In an investigation of the relationship between four variables X_1, X_2, X_3 and X_4, the sample correlation matrix is obtained as given below:

$$R = \begin{bmatrix} 1.000 & 0.715 & -0.526 & 0.807 \\ 0.715 & 1.000 & -0.432 & 0.396 \\ -0.526 & -0.432 & 1.000 & -0.441 \\ 0.8.7 & 0.396 & -0.441 & 1.000 \end{bmatrix}.$$

Now, we want to test the hypothesis H_0 : The variables are pair wise uncorrelated against the alternative hypothesis H_A : Not all pairs of variables are not uncorrelated. It may be seen that $|R| = 0.098$.

Therefore, the value of the test statistics given in Eqn. (64), taking n as suggested by Box(1949), is

$$-2\ln\lambda = \left[64 - \frac{1}{6}(2x4 + 11)\right]\ln(0.098) = 141.30$$

The tabulated value of chi-square with 6 degrees of freedom at 5% level of significance is 12.59 is less than the calculated value. Therefore, we may reject the null hypothesis and conclude that the variables are correlated.

Try an exercise.

E6) Consider the data given in E 4) and test whether ever born children (x_1), number of dead children (x_2), mother's education (x_4) and duration of marriage (x_5) are uncorrelated or not.

Now let us summarize the unit.

22.9 SUMMARY

In this unit we have covered the following points:.

1. We tested the equality of all variances and covariances based on a multivariate random sample from normal population. That is, to investigate the significance of the difference between the variances and covariances.

   Level of Significance (α) and Critical Region is

   $$\chi^2 > \chi^2 a(\alpha) \text{ such that } P\{\chi^2 > \chi^2 a(\alpha)\} == \alpha \text{ and } d = \left\{\frac{1}{2}p(p+1) - 2\right\}$$

   Test Statistic

   $$\chi^2 = -\left[(n-1) - \frac{p(p+1)^2(2p-3)}{6(p-1)(p^2+p-4)}\right]\log\left(\frac{nT}{n-1}\right)$$

   where,

   $$T = \frac{|S|}{(s^2)^p (1-r)^{p-1}(1+(p-1)r)}$$

   The Statistic χ^2 follows χ^2 distribution with $\left\{\frac{1}{2}p(p+1) - 2\right\}$ degrees of freedom.

   Conclusion

   If $\chi^2 \le \chi^2_d(\alpha)$, we conclude that the data do not provide us any evidence against the null hypothesis H_0 and hence it may be accepted at $\alpha\%$ level of significance. Otherwise reject H_0 or accept H_1.

2. We tested for covariance matrix to be equal to a given matrix.

Level of Significance (α) and Critical Region is

$$\chi^2 > \chi^2 a(\alpha) \text{ such that } P\{\chi^2 > \chi^2 a(\alpha)\} = \alpha \text{ and } d = \left\{\frac{1}{2}p(p+1)\right\}$$

The Test Statistic is

$$\chi^2 = np(a - \log g - 1)$$

where $a = \text{tr } \Sigma_0^{-1} S / p$

and $g = |\Sigma_0^{-1} S|^{(1/p)}$

The Statistic χ^2 follows χ^2 distribution with $d = \left\{\frac{1}{2}p(p+1)\right\}$ degrees of freedom.

Conclusion

If $\chi^2 \leq \chi^2 d(s\alpha)$, we conclude that the data do not provide us any evidence against the null hypothesis H_0 and hence it may be accepted at $\alpha\%$ level of significance. Otherwise reject H_0 or accept H_A.

3) We tested for covariance matrix to be equal to a constant multiple of given matrix.

Level of Significance (α) and Critical Region is

$$\chi^2 > \chi^2{}_d(\alpha) \text{ such that } P\{\chi^2 > \chi^2{}_d(\alpha)\} = \alpha \text{ and } d = \left\{\frac{1}{2}(p-1)(p+2)\right\}$$

The Test Statistic is

$$\chi^2 = np \log(a_0 / g_0)$$

where $a_0 = \hat{k}, g_0 = a_0 g$ and $g = |\Sigma_0^{-1} S|^{(1/p)}$

The Statistic χ^2 follows χ^2 distribution with $d = \left\{\frac{1}{2}(p-1)(p+2)\right\}$ degrees of freedom.

Conclusion

If If $\chi^2 \leq \chi^2{}_d(\alpha)$, we conclude that the data do not provide us any evidence against the null hypothesis H_0 and hence it may be accepted at $\alpha\%$ level of significance. Otherwise reject H_0 or accept H_A.

4) We tested for independence of components of variable

Level of Significance (α) and Critical Region is

$$\chi^2 > \chi^2{}_d(\alpha) \text{ such that } P\{\chi^2 > \chi^2{}_d(\alpha)\} = \alpha \text{ and } d = p_1 p_2$$

Test Statistic

$$\chi^2 = -n \ln \prod_{i=1}^{k} (1 - \lambda_i)$$

where $k = \min(p_1, p_2)$ and λ_i's are non zero eigen values of $S_{22}^{-1} S_{21} S_{11}^{-1} S_{12}$.

The statistic χ^2 follows χ^2 distribution with $d = p_1 p_2$ degrees of freedom provided $n - 1 > p_1 + p_2$

Conclusion

If $\chi^2 \le \chi^2_d(\alpha)$, we conclude that the data do not provide us any evidence against the null hypothesis H_0 and hence it may be accepted at $\alpha\%$ level of significance. Otherwise reject H_0 or accept H_A.

5) We tested homogeneity of covariance matrices.
Level of Significance (α) and Critical Region is

$$\chi^2 > \chi^2_d(\alpha) \text{ such that } P\{\chi^2 > \chi^2_d(\alpha)\} = \alpha \text{ and } d = \frac{1}{2}(p(p+1)(k-1))$$

Test Statistic

$$\chi^2 = MC^{-1}$$

where $M = (n-k)\ln|S_u| - \sum_{i=1}^{k}(n_i - 1)\ln|S_{iu}|$

$$S_u = \frac{nS}{n-k} \qquad S_{iu} = \frac{n_i S}{n_i - k}$$

The Statistic χ^2 follows χ^2 distribution with $\frac{1}{2}(p(p+1)(k-1))$ degrees of freedom.

Conclusion

If $\chi^2 \le \chi^2_d(\alpha)$, we conclude that the data do not provide us any evidence against the null hypothesis H_0 and hence it may be accepted at $\alpha\%$ level of significance. Otherwise reject H_0 or accept H_A.

6) We tested for correlation matrix.
Level of Significance (α) and Critical Region is

$$\chi^2 > \chi^2_d(\alpha) \text{ such that } P\{\chi^2 > \chi^2_d(\alpha)\} = \alpha \text{ and } d = \frac{1}{2}(p(p-1))$$

Test Statistic

$$\chi^2 = -n\ln|R|$$

The Statistic χ^2 follows χ^2 distribution with $\frac{1}{2}(p(p-1))$ degrees of freedom.

Conclusion

If $\chi^2 \le \chi^2_d(\alpha)$, we conclude that the data do not provide us any evidence against the null hypothesis H_0 and hence it may be accepted at $\alpha\%$ level of significance. Otherwise reject H_0 or accept H_A.

22.10 SOLUTIONS/ANSWERS

E1) Here $n = 2, p = 2$. Using the formula given in Section 22.2, we may get

$$S = \begin{bmatrix} 55.422 & 6.192 \\ 6.192 & 0.948 \end{bmatrix}$$

$$s^2 = 28.125$$

$$s_r^2 = 6.192$$

$$|S| = 14.1992$$

$$r = 0.8542$$

$$\frac{nT}{n-1} = 0.0661$$

$$\chi^2 = 47.54.$$

Conclusion

Since $x^2 < x^2_{0.05,1}$, H_0 is rejected at 5% level of significance and concluded that body length and body weight of species I are not having same variances and covariances.

E2) $H_0 : \Sigma = \begin{bmatrix} 19.63 & 1.46 \\ 1.46 & 4.31 \end{bmatrix}$ against

$H_A : \Sigma \neq \begin{bmatrix} 19.63 & 1.46 \\ 1.46 & 4.31 \end{bmatrix}$

For the data given on species I, it may be verified that the sample variance covariance matrix $S = \begin{bmatrix} 52.65 & 5.8825 \\ 5.8825 & 0.9013 \end{bmatrix}$

$$\Sigma_0^{-1}S = \begin{bmatrix} 0.0522 & -0.0177 \\ -0.0177 & 0.2380 \end{bmatrix} \begin{bmatrix} 52.65 & 5.8825 \\ 5.8825 & 0.9013 \end{bmatrix} = \begin{bmatrix} 2.6552 & 0.2911 \\ 0.4681 & 0.1104 \end{bmatrix}$$

Hence the calculated value of test statistic can be obtained as

$$\chi^2 = 52.29.$$

Conclusion

Since calculated χ^2 is greater than tabulated χ^2 with 3 degrees of freedom at 5% level of significance therefore we reject the null hypothesis at 5% level of significance and conclude that covariance matrix for length and weight for species I differs from $\begin{bmatrix} 19.63 & 1.46 \\ 1.46 & 4.31 \end{bmatrix}$.

E 3) $H_0 : \Sigma = k\begin{bmatrix} 19.63 & 1.46 \\ 1.46 & 4.31 \end{bmatrix}$, k is unknown

against

$H_A : \Sigma \neq k\begin{bmatrix} 19.63 & 1.46 \\ 1.46 & 4.31 \end{bmatrix}$

For the data given on species I, it may be verified that the sample variance covariance matrix $S = \begin{bmatrix} 52.65 & 5.8825 \\ 5.8825 & 0.9013 \end{bmatrix}$

$$\Sigma_0^{-1}S = \begin{bmatrix} 2.6552 & 0.2911 \\ 0.4681 & 0.1104 \end{bmatrix}, n = 20 \text{ and } p = 2.$$

We may check that $\hat{k} = a_0 = p^{-1} \text{tr}\, \Sigma_0^{-1}S = 1.3773$ and go $= 0.3945$

Hence the calculated value of test statistic can be obtained as

$$\chi^2 = 50.01$$

Conclusion

Since calculated χ^2 is greater than tabulated χ^2 with 2 degrees of freedom at 5% level of significance therefore we reject the null hypothesis at 5% level of significance and conclude that covariance matrix for length and weight for species I is not a constant multiple of $\begin{bmatrix} 19.63 & 1.46 \\ 1.46 & 4.31 \end{bmatrix}$.

E4) $H_0 : \Sigma_{12} = 0$

against

$H_0 : \Sigma_{12} \neq 0$

where $\Sigma_{12} = \begin{bmatrix} \sigma_{14} & \sigma_{15} \\ \sigma_{24} & \sigma_{25} \end{bmatrix}$

For the data given above it may be verified that

$$S = \begin{bmatrix} 5.797 & 1.545 & -6.691 & 8.699 \\ 1.545 & 0.806 & -2.047 & 1.591 \\ -6.691 & -2.047 & 27.897 & -10.497 \\ 8.699 & 1.591 & -10.497 & 20.030 \end{bmatrix}$$

here $n = 64$ and $p_1 = p_2 = 2$.

$$S_{11} = \begin{bmatrix} 5.797 & 1.545 \\ 1.545 & 0.806 \end{bmatrix} S_{12} = \begin{bmatrix} -6.691 & 8.699 \\ -2.047 & 1.591 \end{bmatrix}$$

$$S_{21} = \begin{bmatrix} -6.691 & -2.407 \\ 8.699 & 1.591 \end{bmatrix} S_{22} = \begin{bmatrix} 27.897 & -10.497 \\ -10.497 & 20.030 \end{bmatrix}.$$

Therefore $S_{22}^{-1} S_{21} S_{11}^{-1} S_{12} = \begin{bmatrix} 0.12916 & -0.08974 \\ -0.4093 & 0.67171 \end{bmatrix}$.

The eigen values of this products matrix are $\lambda_1 = 0.7326, \lambda_2 = 0.0683$.

Thus, we have $\chi^2 = -\left[64 - \frac{1}{2}(2+2+3)\right] \ln(1-0.7326)(1-0.0683) = 84.09$

Conclusion

Since calculated χ^2 is greater than tabulated χ^2 with 4 degrees of freedom at 5% level of significance therefore we reject the null hypothesis at 5% level of significance and conclude that ever born children and no of dead children are related with the mother education and duration of marriage.

E5) Here $p = 2\, n_i = 20\, (i = 1,2,3,4)$ and $k = 4$ from the data we have

$$S_{1u} = \begin{bmatrix} 55.422 & 6.192 \\ 6.192 & 0.948 \end{bmatrix} S_{2u} = \begin{bmatrix} 44.366 & 9.928 \\ 9.928 & 4.536 \end{bmatrix}$$

$$S_{3u} = \begin{bmatrix} 15.79 & 1.81 \\ 1.81 & 0.318 \end{bmatrix} S_{4u} = \begin{bmatrix} 50.156 & 5.89 \\ 5.89 & 0.878 \end{bmatrix}$$

$$S_u = \begin{bmatrix} 41.4329 & 5.9553 \\ 5.9553 & 3.7244 \end{bmatrix}, |S_{1u}| = 14.1992$$

$|S_{2u}| = 102.679 |S_{3u}| = 1.7451$

$|S_{4u}| = 9.3449 |S_u| = 118.847$

$M = 171.66, C^{-1} = 0.9525$

Therefore $MC^{-1} = 163.51$

Conclusion

Since calculated MC^{-1} which is approximately distributed as χ^2 with 9 degrees of freedom is greater than tabulated χ^2 with 9 degrees of freedom at 5% level of significance therefore we reject the null hypothesis at 5% level of significance and conclude that the covariance matrixes for the four different species I, II, III and IV are not homogeneous.

E6) The sample correlation matrix can be easily obtained as follows

$$R = \begin{bmatrix} 1.000 & 0.715 & -0.526 & 0.807 \\ & 1.00 & -0.432 & 0.396 \\ & & 1.00 & -0.441 \\ & & & 1.00 \end{bmatrix}$$

$$|R| = 0.098$$

$$\chi^2 = -\left[64 - \frac{1}{6}(2x4 + 11)\ln(0.098)\right] = 141.30$$

Conclusion

Since calculated χ^2 with 6 degrees of freedom is greater than tabulated χ^2 with 6 degrees of freedom at 5% level of significance therefore we reject the null hypothesis at 5% level of significance and conclude that the variables are not uncorrelated.

22.11 PRACTICAL ASSIGNMENT

Session 7

1. Write a program in 'C' language to test for covariance matrix to be equal to a given matrix.