Submitted by-Aditya Gautam

Roll No-12

M.sc. 3rd semester

Date of Assignment-02/11/2020

Date of Submission-12/11/2020

Experiment No-04

Topic- CONDITIONAL DISTRIBUTION, PARTIAL AND MULTIPLE CORRELATION

Problem- The estimate of the mean vector $\underset{\sim}{\mu}$ and the covariance matrix Σ of a 4 variate Normal Distribution based on 36 observations are as follows-

$$\underset{\sim}{\mu}=\begin{pmatrix} 2.70 \\ 35.20 \\ 0.20 \\ 6.12 \end{pmatrix} \qquad \hat{\Sigma}=\begin{pmatrix} 23.804 & & & \\ 4.170 & 864.774 & & \\ 0.535 & 2.2428 & 0.416 & \\ 5.882 & 31.148 & 1.172 & 11.348 \end{pmatrix}$$

(i) Find the estimates of the parameters of the conditional distribution of (X_1,X_2) given (X_3,X_4) and that of the conditional distribution of (X_3,X_4) given (X_1,X_2) .

(ii) Find the sample multiple correlation coefficient of X_1 on (X_2,X_3,X_4) and hence ,test the hypothesis $H_0:\rho_{1.234}=0$ at 5% level.

(iii) Find the sample correlation coefficient of X_3 and X_4 (eliminating the effects of X_1 and X_2) and test for $H_0:\rho_{34.12}=0$ at 5% level.

Theory-

(i) We first find the parameters of the conditional distribution of (X_1, X_2) given (X_3, X_4) by means of the following partitioning of μ and $\widehat{\Sigma}$

$$\underset{\sim}{X}=\begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_3 \\ X_4 \end{bmatrix}=\begin{bmatrix} \underset{\sim}{X}^{(1)} \\ \underset{\sim}{X}^{(2)} \end{bmatrix} \qquad \widehat{\underset{\sim}{\mu}}=\begin{bmatrix} \widehat{\mu_1} \\ \widehat{\mu_2} \\ \cdots \\ \widehat{\mu_3} \\ \widehat{\mu_4} \end{bmatrix}=\begin{bmatrix} \underset{\sim}{\mu}_1^{(1)} \\ \underset{\sim}{\mu}_2^{(2)} \end{bmatrix}$$

$$\widehat{\Sigma} = \begin{pmatrix} \widehat{\sigma_{11}} & \widehat{\sigma_{12}} & \vdots & \widehat{\sigma_{13}} & \widehat{\sigma_{14}} \\ \widehat{\sigma_{21}} & \widehat{\sigma_{22}} & \vdots & \widehat{\sigma_{23}} & \widehat{\sigma_{24}} \\ \cdots & \cdots \quad \cdots & \vdots & \cdots \quad \cdots & \cdots \\ \widehat{\sigma_{31}} & \widehat{\sigma_{32}} & \vdots & \widehat{\sigma_{33}} & \widehat{\sigma_{34}} \\ \widehat{\sigma_{41}} & \widehat{\sigma_{42}} & \vdots & \widehat{\sigma_{43}} & \widehat{\sigma_{44}} \end{pmatrix}=\begin{pmatrix} \widehat{\Sigma_{11}} & \widehat{\Sigma_{12}} \\ \widehat{\Sigma_{21}} & \widehat{\Sigma_{22}} \end{pmatrix}$$

The estimates of the parameter of the conditional distribution of (X_1, X_2) given (X_3, X_4) are given by -

$\hat{\mu}=\mu^{(1)}+\widehat{\beta_1}(x^{(2)} - \underset{\sim}{\mu}^{(2)})$

Where, $\widehat{\beta_1}=\widehat{\Sigma_{12}}\widehat{\Sigma_{22}}^{-1}$ $\qquad$ $\widehat{\Sigma_{11.2}} = \widehat{\Sigma_{11}} - \widehat{\Sigma_{12}}\, \widehat{\Sigma}^{-1}_{22}\, \widehat{\Sigma}_{21}$

The parameter of the conditional distribution of (X_3,X_4) given (X_1,X_2) then we found by means of the following partitioned vector is as follows:-

$$\underset{\sim}{X}=\begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_3 \\ X_4 \end{bmatrix}=\begin{bmatrix} \underset{\sim}{X}^{(1)} \\ \underset{\sim}{X}^{(2)} \end{bmatrix} \qquad \widehat{\underset{\sim}{\mu}}=\begin{bmatrix} \widehat{\mu_3} \\ \widehat{\mu_4} \\ \cdots \\ \widehat{\mu_1} \\ \widehat{\mu_2} \end{bmatrix}=\begin{bmatrix} \underset{\sim}{\mu}_1^{(1)} \\ \underset{\sim}{\mu}_2^{(2)} \end{bmatrix}$$

$$\hat{\Sigma} = \begin{pmatrix} \widehat{\sigma_{33}} & \widehat{\sigma_{34}} & & \vdots & \widehat{\sigma_{31}} & & \widehat{\sigma_{32}} \\ \widehat{\sigma_{43}} & \widehat{\sigma_{44}} & & \vdots & \widehat{\sigma_{41}} & & \widehat{\sigma_{42}} \\ \cdots & \cdots & \cdots & \vdots & \cdots & \cdots & \cdots \\ \widehat{\sigma_{13}} & \widehat{\sigma_{14}} & & \vdots & \widehat{\sigma_{11}} & & \widehat{\sigma_{12}} \\ \widehat{\sigma_{23}} & \widehat{\sigma_{24}} & & \vdots & \widehat{\sigma_{21}} & & \widehat{\sigma_{22}} \end{pmatrix} = \begin{pmatrix} \widehat{\Sigma_{11}} & \widehat{\Sigma_{12}} \\ \widehat{\Sigma_{21}} & \widehat{\Sigma_{22}} \end{pmatrix}$$

The estimates of the parameter of the conditional distribution of (X_3, X_4) given (X_1, X_2) are given by-

$\hat{\mu}_{34.12}=\mu^{(1)}+\hat{\beta}_1\ (x^{(2)}-\underset{\sim}{\mu}^{(2)})$

Where, $\hat{\beta}_1=\hat{\Sigma}_{12}\hat{\Sigma}^{-1}_{22}$

$\hat{\Sigma}_{22.1}=\hat{\Sigma}_{22}-\hat{\Sigma}_{21}\hat{\Sigma}^{-1}_{11}\hat{\Sigma}_{12}$

(ii) Here, we first partitioned $\underset{\sim}{X}$, $\underset{\sim}{\hat{\mu}}$ and $\hat{\Sigma}$ as follows: -

$$\underset{\sim}{X}=\begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_3 \\ X_4 \end{bmatrix}=\begin{bmatrix} \underset{\sim}{X}^{(1)} \\ \underset{\sim}{X}^{(2)} \end{bmatrix} \qquad \underset{\sim}{\hat{\mu}}=\begin{bmatrix} \widehat{\mu_1} \\ \widehat{\mu_2} \\ \cdots \\ \widehat{\mu_3} \\ \widehat{\mu_4} \end{bmatrix}=\begin{bmatrix} \underset{\sim}{\mu}_1^{(1)} \\ \underset{\sim}{\mu}_2^{(2)} \end{bmatrix}$$

$$\hat{\Sigma} = \begin{pmatrix} \widehat{\sigma_{11}} & \widehat{\sigma_{12}} & & \vdots & \widehat{\sigma_{13}} & & \widehat{\sigma_{14}} \\ \widehat{\sigma_{21}} & \widehat{\sigma_{22}} & & \vdots & \widehat{\sigma_{23}} & & \widehat{\sigma_{24}} \\ \cdots & \cdots & \cdots & \vdots & \cdots & \cdots & \cdots \\ \widehat{\sigma_{31}} & \widehat{\sigma_{32}} & & \vdots & \widehat{\sigma_{33}} & & \widehat{\sigma_{34}} \\ \widehat{\sigma_{41}} & \widehat{\sigma_{42}} & & \vdots & \widehat{\sigma_{43}} & & \widehat{\sigma_{44}} \end{pmatrix} = \begin{pmatrix} \widehat{\Sigma_{11}} & \widehat{\Sigma_{12}} \\ \widehat{\Sigma_{21}} & \widehat{\Sigma_{22}} \end{pmatrix}$$

The multiple correlation coefficient between X_1 on (X_2, X_3, X_4) denoted by $R_{1.234}$ is given by-

$R^2_{1.234}=\frac{\widehat{\Sigma_{12}}\widehat{\Sigma_{22}^{-1}}\widehat{\Sigma_{21}}}{\widehat{\Sigma_{11}}}$

To test the hypothesis H_0:$\rho_{1.234}$=0(X_1 is independent of (X_2, X_3, X_4). We use the following to test the statistics under H_0

$\frac{R^2_{1.234}}{1-R^2_{1.234}} \times (\frac{n-p-1}{p})\sim F(p,n-p-1)$

Where, 'n' is the number of observations.

The calculated value is compared with the tabulated value and conclusions are done accordingly.

(iii) Let, $\hat{\Sigma}_{22.1}=\begin{pmatrix} a_1 & a_2 \\ a_2 & a_3 \end{pmatrix}$

Then $r_{11.2}=\frac{cov(X_{3.12},X_{4.12})}{\sqrt{var(X_{3.12})var(X_{4.12})}}$

$=\frac{a_2}{\sqrt{a_1 \times a_3}}$

Test the null hypothesis $H_0:\rho_{1.234}=0$, then test statistic under H_0 is

$Z=\frac{Z-\xi}{\sqrt{\frac{1}{n-3}}} \sim N(0,1)$

Where, $Z=\frac{1}{2}log_e(\frac{1+r_{34.12}}{1-r_{34.12}})$

$\xi=\frac{1}{2}log_e(\frac{1+r_{34.12}}{1-r_{34.12}})$

$=\frac{1}{2}log_e 1=0$

We compare this value with the tabulated value of Z at 5% level and draw the conclusions accordingly.

Calculation-

The R-Programming for obtaining the solution.

```
s=array(c(23.804,4.170,0.535,5.882,4.170,864.774,2.2428,31.148,0.535,2.2428,0.416,1.172,5.882,31.148,1.172,11.348),dim=c(4,4))

s

s11=array(c(23.804,4.170,4.170,864.774),dim=c(2,2))

s11

s12=array(c(0.535,2.2428,5.882,31.148),dim=c(2,2))

s12

s21=t(s12)

s21

s22=array(c(0.416,1.172,1.172,11.348),dim=c(2,2))

B1=s12%*%solve(s22)

B1

s11_2=s11-(B1%*%s21)

s11_2

B2=s21%*%solve(s11)

B2

s22_1=s22-(B2%*%s12)

s22_1

a1=s22_1[1,1]

a1

a2=s22_1[1,2]

a2

a3=s22_1[2,2]
```

```
a3
r34_12=a2/sqrt(a1*a3)
r34_12
z=1/2*log((1+r34_12)/(1-r34_12))
z
n=36
p=4
Z=z/sqrt(1/(n-3))
Z
Z_tab=1.96
w11=23.804
w12=array(c(4.170,0.535,5.882),dim=c(1,3))
w12
w21=t(w12)
w21
w22=array(c(864.774,2.2428,31.148,2.2428,0.416,1.172,31.148,1.172,11.348),dim=c(3,3))
w22
R1_234=sqrt((w12%*%solve(w22)%*%w21)/w11)
R1_234
t=(R1_234^2/(1-(R1_234^2)))*((n-p-1)/p)
t
t_tab=qf(0.95,4,31,0)
t_tab
```

Result and calculations-

(i) The estimates of the parameters of the conditional distribution of (X_1, X_2) given (X_3, X_4) is

$$\widehat{\mu}_{12..34} = \begin{pmatrix} 2.70 \\ 35.20 \end{pmatrix} + \begin{pmatrix} -0.2457361 & 0.5437084 \\ -3.3025258 & 3.0858795 \end{pmatrix} \begin{pmatrix} x_3 & -0.20 \\ x_4 & -6.12 \end{pmatrix}$$

The estimates of the parameters of the conditional distribution of (X_3, X_4) given (X_1, X_2) is

$$\widehat{\mu}_{34.12} = \begin{pmatrix} 0.20 \\ 6.12 \end{pmatrix} + \begin{pmatrix} 0.0220395 & 0.002487234 \\ 0.2409951 & 0.034856564 \end{pmatrix} \begin{pmatrix} x_1 & -2.70 \\ x_2 & -6.12 \end{pmatrix}$$

(ii) The multiple correlation coefficient of X_1 on (X_2, X_3, X_4) denoted by $R_{1.234}$ is given by

$$R_{1.234} = 0.3700054$$

Since, the calculated value is < the tabulated value of F= 2.678667 at 5% loss, so we accept the null hypothesis and conclude that X_1 is independent of (X_2, X_3, X_4).

(iii) The partial correlation coefficient between X_3 and X_4 (after eliminating the joint effect of X_1 and X_2 from each of them) is $r_{34.12}$= 0.5138664

Since, the calculated value of Z= 3.262736 > the tabulated value, therefore we reject our null hypothesis H_0 and conclude that there exist partial correlation coefficient between X_3 and X_4.