Submitted by-Aditya Gautam

Roll No-12

M.sc. 3rd semester

Date of Assignment-10/12/2020

Date of Submission-17/12/2020

Experiment No-09

Topic- MAHALANOBIS D^2 STATISTIC.

Problem- The following data shows the marks actually obtained by 10 students and the expected marks of 12 students in an examination. Test the null hypothesis of equality of the marks actually obtained and the expected mark assuming Tri-variate normal distribution.

Population-I(marks scored)			
Student	I $(X_1)_1$	II $(X_2)_1$	III$(X_3)_1$
1	65	67	67
2	66	76	89
3	63	64	66
4	60	70	69
5	60	60	80
6	62	66	65
7	71	57	69
8	60	68	70
9	68	67	70
10	71	84	87

Population-II(expected marks)			
Student	$I(X_1)_1$	$II(X_2)_1$	$III(X_3)_1$
1	75	70	72
2	56	65	68
3	71	75	90
4	66	72	80
5	72	75	80
6	64	69	71
7	59	65	74
8	66	65	76
9	71	54	75
10	68	70	72
11	64	72	73
12	65	86	90

Theory-

Let $\begin{pmatrix} (X_{11})_1 \dots\dots & (X_{p1})_1 \\ \dots\dots\dots & \\ \dots\dots & \\ (X_{1n_1})_1 \dots\dots & (X_{pn_1})_1 \end{pmatrix}$ & $\begin{pmatrix} (X_{11})_2 \dots\dots & (X_{p1})_2 \\ \dots\dots\dots & \\ \dots\dots & \\ (X_{1n_2})_2 \dots\dots & (X_{pn_2})_2 \end{pmatrix}$ be

the two samples of sizes n_1 *and* n_2 drawn from the multivariate normal populations $N_1\left(\underset{\sim}{\mu}_1, \Sigma\right)$ & $N_2\left(\underset{\sim}{\mu}_2, \Sigma\right)$. Here it is assumed that both the populations have the same variance covariance matrix. The Mahalanobis D^2 statistic is defined as

$D^2 = (\underset{\sim}{\bar{X}}_1, \underset{\sim}{\bar{X}}_2)' S^{-1}{}_{pooled} (\underset{\sim}{\bar{X}}_1, \underset{\sim}{\bar{X}}_2)$ *where*

$$\underset{\sim}{\bar{X}}_1 = \left((\bar{X}_1)_1, (\bar{X}_2)_1, \dots., \left(\bar{X}_p\right)_1\right)'; (\bar{X}_i)_1 = \frac{1}{n_1}\sum_{j=1}^{n_1}(X_{ij})_1$$

$\underset{\sim}{\bar{X}}_2 = \left((\bar{X}_1)_2, (\bar{X}_2)_2, \dots., \left(\bar{X}_p\right)_2\right)'; (\bar{X}_i)_2 = \frac{1}{n_1}\sum_{j=1}^{n_2}(X_{ij})_2$, $S_{pooled} = \frac{(n_1-1)S_1}{(n_1+n_2-2)} + \frac{(n_2-1)S_2}{(n_1+n_2-2)}$

$S_1 = ((S_{ij})_1); (S_{ij})_1 = \frac{1}{n_1-1}\sum_{k=1}^{n_1}\{(X_{ik})_1 - (\bar{X}_i)_1\}\{(X_{jk})_1 - (\bar{X}_j)_1\}$

$S_2 = ((S_{ij})_2); (S_{ij})_2 = \frac{1}{n_2-1}\sum_{k=1}^{n_2}\{(X_{ik})_2 - (\bar{X}_i)_2\}\{(X_{jk})_2 - (\bar{X}_j)_2\}$

Here, we are to test the hypothesis $H_0: \underset{\sim}{\boldsymbol{\mu}}_1 = \underset{\sim}{\boldsymbol{\mu}}_2$ i.e. the two population means are equal against $H_1: \underset{\sim}{\mu}_1 \neq \underset{\sim}{\mu}_2$.

The Mahalanobis-D^2 test is $\frac{n_1+n_2-p-1}{(n_1+n_2-2)p}\left(\frac{n_1 n_2}{n_1+n_2}\right)D^2 \sim F_{p, n_1+n_2-p-1}$.

The conclusions are drawn accordingly.

Calculation-

The R-programming to obtain the solution for the given problem-

```
x1=c(65,66,63,60,60,62,71,60,68,71,67,76,64,70,60,66,57,68,67,84,67,89,66,69,80,65,69,70,70,
87)

x1

dim(x1)=c(10,3)

dim(x1)

x2=c(75,56,71,66,72,64,59,66,71,68,64,65,70,65,75,72,75,69,65,65,54,70,72,86,72,68,90,80,80,
71,74,76,75,72,73,90)

x2

dim(x2)=c(12,3)

dim(x2)

m1=mat.or.vec(3,1)

m2=mat.or.vec(3,1)

for(i in 1:3){

m1[i]=mean(x1[,i])

m2[i]=mean(x2[,i])}

mean=array(c(m1-m2),dim=c(3,1))

mean

n1=10

n2=12

p=3

var11=mat.or.vec(3,1)

var12=mat.or.vec(3,1)

var21=mat.or.vec(3,1)

var22=mat.or.vec(3,1)
```

```
var31=mat.or.vec(3,1)
var32=mat.or.vec(3,1)
for(i in 1:3){
var11[i]=cov(x1[,1],x1[,i])*((n1-1)/(n1+n2-2))
var12[i]=cov(x2[,1],x2[,i])*((n2-1)/(n1+n2-2))}
for(i in 1:3){
var21[i]=cov(x1[,2],x1[,i])*((n1-1)/(n1+n2-2))
var22[i]=cov(x2[,2],x2[,i])*((n2-1)/(n1+n2-2))}
for(i in 1:3){
var31[i]=cov(x1[,2],x1[,i])*((n1-1)/(n1+n2-2))
var32[i]=cov(x2[,2],x2[,i])*((n2-1)/(n1+n2-2))}
s1=c(var11,var21,var31)
s1
dim(s1)=c(3,3)
dim(s1)
s2=c(var12,var22,var32)
s2
dim(s2)=c(3,3)
dim(s2)
s_p=s1+s2
s_p
d2=t(mean)%*%solve(s_p)%*%mean
d2
cal=((n1+n2-p-1)/(p*(n1+n2-2)))*((n1*n2)/(n1+n2))*d2
cal
```

tab=qf(0.95,3,18,0)

tab

Conclusion-

The MahalanobisD^2 -statistic is 0.259089. Since the calculated value of F(i.e0.4239638) is less than the tabulated value of F(i.e 3.159908), hence we accept our null hypothesis at 5% level of significance and conclude that the two population means are equal.