Submitted by-Aditya Gautam

Roll No-12

M.sc. 3rd semester

Date of Assignment-10/11/2020

Date of Submission-16/11/2020

Experiment No-05

Topic- FITTING OF MULTIPLE REGRESSION MODEL AND COMPUTATION OF MULTIPLE CORRELATION COEFFICIENT.

Problem- A study was carried out in the class to see if in a particular subject, the performance of the students in the 3 class tests affect the performance in the final examination. The marks scored out of 100 in the final examination and out of 15 in each of the 3 class tests in the subject under consideration by 15 students of the class is given below.

Sl. No. of students	**Final exam marks(Y)**	**Marks in 1st class test(X_1)**	**Marks in 2nd class test(X_2)**	**Marks in 3rd class test(X_3)**
1.	87	45	41	49
2.	97	34	41	38
3.	81	40	35	43
4.	75	37	35	40
5.	69	31	36	39
6.	78	42	31	37
7.	91	48	47	48
8.	83	41	39	43
9	93	49	50	46
10.	72	38	36	42
11.	80	38	40	38
12.	89	41	47	38
13.	81	39	40	37
14.	78	36	33	39
15.	59	29	22	30

(i)Fit a multiple linear regression model to the given data.

(ii)Compare the multiple correlation coefficient of Y on (X_1,X_2,X_3) and hence, test if it is significantly different from 0.

Theory- (i) In matrix notation, the multiple linear regression model is given by

$\tilde{Y}=\tilde{X}\,\tilde{\beta} + \tilde{\varepsilon}$

Where, $\tilde{Y}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $\tilde{\beta}=\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$

$$\tilde{X}=\begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \ldots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \ldots & X_{np} \end{pmatrix}$$

$\tilde{\varepsilon}=(\varepsilon_1, \varepsilon_2, \ldots \ldots \ldots \ldots \ldots \varepsilon_n)$

$Y_i=\beta_1 + \beta_2 X_{i1} + \beta_3 X_{i2} + \beta_4 X_{i3}$

$Y_1=\beta_1 + \beta_2 X_{11} + \beta_3 X_{12} + \beta_4 X_{13}$

$Y_2=\beta_1 + \beta_2 X_{21} + \beta_3 X_{22} + \beta_4 X_{23}$

$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$

$Y_n=\beta_1 + \beta_2 X_{n1} + \beta_3 X_{n2} + \beta_4 X_{n3}$

The least square estimate of $\tilde{\beta}$ is given by

$\underset{\sim}{\tilde{\beta}} = (X'X)^{-1}(X'Y)$

and thus the fitted multiple linear regression model becomes,

$Y=\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \ldots\ldots \widehat{\beta_p} X_p$

(ii) We first find the sample covariance matrix of X and partitioned it as shown below:

$$\underset{\sim}{X} = \begin{pmatrix} Y \\ \ldots \\ X_1 \\ X_2 \\ X_3 \end{pmatrix} \qquad \hat{\Sigma} = S = \begin{bmatrix} \sigma_y^2 & \vdots & \sigma_{y1} & \sigma_{y2} & \sigma_{y3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{y1} & \vdots & \sigma_{11}^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{y2} & \vdots & \sigma_{21} & \sigma_{22}^2 & \sigma_{23} \\ \sigma_{y3} & \vdots & \sigma_{31} & \sigma_{32} & \sigma_{33}^2 \end{bmatrix} = \begin{bmatrix} \hat{\Sigma}_{11} & \widehat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{bmatrix}$$

The multiple correlation coefficient between Y and (X_1, X_2, X_3) is given by

$R_{y.123}=R^2=\frac{\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}}{\hat{\Sigma}_{11}}$

The null to be tested here is $H_0:\rho_{4.123}=0$

(The population multiple correlation coefficient is not significantly different from zero)

Under H_0 the test statistic is

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

The calculated value of F is compared to the tabulated values and conclusions are drawn according

Calculation-

The R-Programming for obtaining the solution.

```
Y=c(87,97,81,75,69,78,91,83,93,72,80,89,81,78,59)

X1=c(45,34,40,37,31,42,48,41,49,38,38,41,39,36,29)

X2=c(41,41,35,35,36,31,47,39,50,36,40,47,40,33,32)

X3=c(49,38,43,40,39,37,48,43,46,42,38,38,37,39,30)

X0=rep(1,times=15)

X0

X=array(c(X0,X1,X2,X3),dim=c(15,4))

X

X_X=t(X)%*%X

X_X

X_Y=t(X)%*%Y

X_Y

B=solve(X_X)%*%X_Y

B

Z=array(c(Y,X1,X2,X3),dim=c(15,4))

Z
```

```
S=cov(Z)

S

S11=S[1,1]

S11

S12=array(c(S[1,2],S[1,3],S[1,4]),dim=c(1,3))

S12

S21=t(S12)

S21

S22=array(c(S[2,2],S[3,2],S[4,2],S[2,3],S[3,3],S[4,3],S[2,4],S[3,4],S[4,4]),dim=c(3,3))

S22

R=sqrt((S12%*%solve(S22)%*%S21)/S11)

R

R2=R^2

R2

n=15

p=4

F_cal=(R2/(1-R2))*((n-p-1)/p)

F_cal

F_tab=qf(0.95,n,n-p-1)

F_tab
```

Conclusion-

(i) The fitted multiple linear regression model is -
$\underline{Y=19.1718726+0.4107971X_1+0.9546416X_2+0.2097482X_3}$

(ii) The multiple correlation coefficient between Y and (X_1,X_2X_3)is given by

$R_{Y.123}$= 0.7985956

Since, The calculated value of F at (p,n-p-1) = (4,10) d.f is 4.401406 which is greater than the tabulated value of F= 2.845017 at 5% level of significance, so we reject the null hypothesis and conclude that the population multiple correlation is not significantly different from zero(0).