Submitted by-Aditya Gautam

Roll No-12

M.sc. 3rd semester

Date of Assignment-21/10/2020

Date of Submission-31/10/2020

**Experiment No-03**

**Topic-** FITTING OF MULTIPLE LINEAR REGRESSION MODEL.

**Problem-** The following arrangement of the sum of squares and sum of product corrected for mean, each being based on 28 observations.



1. Fit a regression model of the form



1. Test for the significance of the multiple correlation coefficient of Y on (X1, X2, X3).

Also construct the 95% confidence interval for each of the regression coefficient.

**Theory-**

1. Suppose we are to fit a regression model of the form-

 where Y is the dependent variables, Xi’ s are independent variables (i=1,2,3, ………, p) and ’s are the regression coefficient (i=1,2,3,…………., p). Let us further suppose that there are ‘n’ observations under each of the ‘p-variable’ and Y. Therefore, for the ith observation , we have-



Here it is assume that-



Also, Normal distribution with mean 0 and variance 

Taking sum over all the ‘n’ observation on equation (1) and dividing by ‘n’ we get—



 ---------------------(2)

Subtracting eqn (2) from each of the ‘n’ equation in (1), we get-



 ----------------(3)

Where  denote deviation of the ith observation under kth variable from its mean,

k=1,2,3,………………p and i=1,2,3,……………..n

The ‘n’ equations in no (3) may be written in the matrix form as-



Where,

 ,  ,  , 

The least square estimates of  is given by-



Where,

 , and 

The required multiple linear regression model to be fitted is -



1. Here the null hypothesis to be tested is -

H0: The multiple correlation coefficient of Y on (X1, X2, X3) is zero.

Under H0 the test statistic is-



Where, R is the multiple correlation coefficient of a variable with the other ‘p’ variables in a random sample of size ‘n’ from a (p+1) variate distribution.



The calculated value is compared with the tabulated value and conclusion are drawn accordingly.

**Confidence Interval Contraction-**

Suppose in a regression model  , X has full rank ‘p’ and  is distributed as . The simultaneous 100(1-α)% confidence interval for  is given by -



Where  is the diagonal element of  corresponding to  and  and  is the upper 100(1-α)th percentile of the F distribution with p and (n-p) d.f.

The program for finding the solution to the given problem is-

 , and 

**R Program-**

XX=array(c(10.458,0.238,7.032,0.238,1.316,-0.268,7.032,-0.268,16.047),dim=c(3,3))

XX

XY=array(c(1.379,2.198,0.368),dim=c(3,1))

XY

XX\_inv=solve(XX)

B=(XX\_inv)%\*%(XY)

B

den=4.244

num=den-(t(B)%\*%(XX)%\*%B)

n=28

p=3

R\_sqr=1-((num/den)\*((n-1)/(n-p)))

R\_sqr

F\_cal=( R\_sqr/(1- R\_sqr))\*((n-p-1)/p)

F\_cal

F\_tab=qf(1-0.05,3,24)

F\_tab

EE=num

S\_sqr=EE/(n-p)

V=S\_sqr[1,1]\*(XX\_inv)

V

V\_B=array(c(sqrt(V[1,1]), sqrt(V[2,2]), sqrt(V[3,3])),dim=c(3,1))

T2=sqrt(p\*qf(0.05,3,25))

LCL=B-(V\_B\*T2)

LCL

UCL= B+(V\_B\*T2)

UCL

#We write the following program to get the pair (LCL, UCL) for Beta1, Beta2 and Beta3

CONF\_INTRVL=mat.or.vec(3,2)

for (i in 1:3){

CONF\_INTRVL[i,1]=c(LCL[i])

CONF\_INTRVL[i,2]=c(UCL[i])}

CONF\_INTRVL

**#Result from the R-Programming-**

#F\_cal=57.7116

#F\_tab=3.008787

#THE VALUE OF REGRESSION COEFFICIENTS ARE-

#Beta1=0.08521572

#Beta2=1.65750420

#Beta3=0.01327190

#AND CONFIDENCE INTERVAL FOR Beta1, Beta2 AND Beta3 IS GIVEN BY-

# 0.05499460 0.11543684

# 1.58583258 1.72917582

# -0.01111644 0.03766024

**Result and Conclusion-**

1. The required multiple linear regression model to be fitted is -



1. Since the calculated value of F (i.e. cal F=57.7116) is greater than the tabulated value of F ( i.e. tab F=3.008787). Therefore we reject our null hypothesis & we conclude that the multiple correlation coefficient of Y on (X1, X2, X3) is not equal to zero.

The 95% confidence interval for βi are given below-

For β1 is = (0.05499460, 0.11543684)

For β2 is = (1.58583258, 1.72917582)

For β3 is = (-0.01111644, 0.03766024)