LAB NO 15(REGRESSION)

Q.1) A computer manager needs to know how efficiency of her new computer program depends on the size of incoming data and how many tables are used to arrange each data set. Efficiency will be measured the number of processed requests per hour. Applying the program to data set of different sizes and number of tables are used, she gets the following results.

Processed request, Y	Data size,(gigabyte),X1	Number of tables,X2
16	15	1
26	10	2
17	10	10
41	8	10
50	7	20
55	7	20
40	6	4

a. Write the regression equation for processed request.
b. Interpret the parameters of the regression model.
c. What percentage of variation on processed request is explained by two independent variables.
d. Computer standard error of the estimate.
e. Also, computer adjusted R square.
f. Test the significance of each of the regression coefficient.
g. Test the overall goodness of fit of the model.

Solution:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.887368
R Square	0.787421
Adjusted R Square	0.681132
Standard Error	8.784418
Observations	7

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1143.336	571.668	7.408288	0.04519
Residual	4	308.664	77.166		
Total	6	1452			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	52.73476	16.86809	3.126302	0.035311
X Variable 1	-2.87174	1.436068	-1.99973	0.116153
X Variable 2	0.84741	0.551494	1.536572	0.199207

a. Let the regression equation be Y=a+b1X1+b2X2
   From coefficient table,
   Y=51.56-2.87X1+0.84X2

b. Here a=52.76, if X1 and X2 becomes 0 then efficiency becomes 51.56.
   b1=-2.87 i.e. if we increase the data size by one unit then efficiency decreases by 2.62 unit keeping the effect of number of tables as constant.
   b2=0.84 i.e. if we increase the data size by one unit then number of tables increases by 0.84 unit keeping the effect of efficiency as constant.

c. R2 = 0.78 i.e. 78% of total variation on processed request is explained by two independent variables.

d. Se = 8.78 i.e. the average deviation of observation from the fitted regression line is 8.78.

e. Adjusted R2 = 0.68

f. Test for B1

   Hypothesis:

   H0: Regression coefficient is not significant.
   H1: Regression coefficient is significant.

   Level of significance: alpha = 5%

   Test statistics:

   t=1.99

   P-value:

   P-value=0.116

   Decision:

   Since, P-value>Alpha so we do not reject H0.

   Hence, we conclude that regression coefficient is not significant.

Test for B2

Hypothesis:

H0: Regression coefficient is not significant.
H1: Regression coefficient is significant.

Level of significance: alpha = 5%

Test statistics:

t=1.53

P-value:

P-value=0.199

Decision:

Since, P-value>Alpha so we do reject H0.

Hence, we conclude that regression coefficient is not significant.

g. Test for regression model

Hypothesis:

H0: Regression model is not significant.
H1: Regression model is significant.

Level of significance: alpha = 5%

Test statistics:

F=7.408

P-value:

P-value=0.045

Decision:

Since, P-value>Alpha so we reject H0.

Hence, we conclude that regression model is significant.

Q.2) It was reported somewhere that children whenever playing the game on computer, they use the computer very roughly which may reduce the lifetime of a computer. The random-access memory (RAM) of a computer also plays a crucial role in the lifetime of a computer. A researcher wanted to examine how the lifetime of a personal computer that is used by children is affected by the time (in hours) spent by the children per day playing games and the available random-access memory (RAM) measured in megabytes (MB) of a used computer. The data is provided in the following table.

Lifetime(years)	Play time(hours)/day	RAM in Mb
5	2	8
1	8	2
7	1	6
2	5	3
3	6	2
4	3	4
6	2	7

a) Write the estimated regression equation for the lifetime.
b) Interpret the parameters of the regression model.
c) What percentage of variation in lifetime is explained by two independent variables?
d) Compute the standard error of the estimate.
e) Also compute adjusted R square.
f) Test the significance of each of the regression coefficients.
g) Test the overall goodness of fit of the model.

Solution:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.939861914
R Square	0.883340416
Adjusted R Square	0.825010625
Standard Error	0.903668681
Observations	7

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	24.73353166	12.36676583	15.143898	0.013609458
Residual	4	3.266468338	0.816617085		
Total	6	28			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
Intercept	6.961325967	2.481625648	2.805147493	0.048556218	0.071228582
X Variable 1	-0.785380365	0.29455627	-2.666316917	0.056020402	-1.603199679
X Variable 2	0.014874628	0.307243511	0.048413156	0.963707852	-0.838170114

a. Let the regression equation be: $Y=a+b1x1+b2x2$

   From the coefficient table,
   $Y= 6.91 – 0.78x1+0.01x2$

b. Here a= 6.91 i.e. the lifetime will be 6.91 if we keep both independent variables zero
   B1=-0.78 i.e. if we increase the value of playtime by one unit then the lifetime will be decreased by 0.78 keeping the effect of RAM constant
   B2= 0.01 i.e. if we increase the value of RAM by one unit then the lifetime will be increased by 0.01 keeping the effect of play time constant.

c. R Square = 0.88 i.e. 88% of total deviation on lifetime is explained by two independent variables.

d. Standard error = 0.90 i.e. the average deviation from the fitting regression line is 0.90

e. Adjusted R square = 0.82

f.

Test for B1,

Hypothesis:

H0: The regression coefficient isn't significant.
H1: The regression coefficient is significant.

Level of significance: Alpha = 5%

Test statistics:

T=2.66
P value =0.056

Decision, Since the P-value is greater than alpha so we don't reject H0.
Hence, we conclude that the regression coefficient is not significant.

Test for B2,

Hypothesis:

H0: The regression coefficient isn't significant.
H1: The regression coefficient is significant.

Level of significance: Alpha = 5%

Test statistics:

T=0.04
P value =0.96

Decision, Since the p-value is greater than the alpha so we accept H0.
Hence, we conclude that the regression coefficient isn’t significant.

Test for regression model

Hypothesis:

H0: The regression model isn’t significant.
H1: The regression model is significant.

Level of significance: Alpha = 5%

Test statistics:

F =15.14
P value =0.01

Decision, since the p-value is smaller than alpha we don’t reject H0.
Hence, we conclude that the regression model isn’t significant.