Regression

Lab no 15:

A computer manager needs to know how the 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 by the number of processed requests per hour. Applying the program to data sets of different sizes and number of tables, she gets the following results.

|  |  |  |
| --- | --- | --- |
| Processed requests Y | Data size, (GB), X1 | Number of tables, X2 |
| 16 | 15 | 1 |
| 26 | 10 | 10 |
| 41 | 8 | 10 |
| 50 | 7 | 20 |
| 55 | 7 | 20 |
| 40 | 6 | 4 |

1. Write the regression equation for the processed request.
2. Interpret the parameters of the regression model.
3. What percentage of variation on processed requests is explained by two independent variables?
4. Compute the standard error of the estimate.
5. Also compute adjusted R square.
6. Test the significance of each of the regression coefficients.
7. Test the overall goodness of fit of the model.

Solution  


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

From the coefficient table,

1. Y=51.56-2.62x1+0.89x2 which is a required equation.
2. Here,

a= 51.56 i.e. if x1 and x2 become zero then efficiency becomes 51.56.

b1=-2.62 i.e. if we increase the data size by one unit then efficiency decreases by 2.62 units keeping the effect of several tables as constant.

B2=0.89 i.e. if we increase the number of tables by one unit then efficiency increases by 0.89 keeping the effect of data size constant.

1. R2 = 0.91 i.e. 91% of total variation on processed requests is explained by two independent variables.
2. Standard error = 5.65 i.e. the average deviation of observation from the fitted regression line is 5.65.
3. Adjusted R2= 0.85
4. Test for B1,

Hypothesis:

H0: The regression coefficient isn’t significant.

H1: The regression coefficient is significant.

Alpha = 5%

Test statistics:

T=2.81

P value =0.066

The 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.

Alpha = 5%

Test statistics:

T=2.28

P value =0.01

Decision, since the p value is less than alpha we reject H0.

Hence, we conclude that the regression coefficient is significant.

1. Test for regression model

Hypothesis:

H0: The regression model isn’t significant.

H1: regression model is significant.

Alpha = 5%

Test statistics:

F =15.31

P value =0.026

The decision, since the p value is greater than alpha so we reject H0.

Hence, we conclude that the regression model is significant.

Lab no. 16

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 |

1. Write the estimated regression equation for the lifetime.
2. Interpret the parameters of the regression model.
3. What percentage of variation in lifetime is explained by two independent variables?
4. Compute the standard error of the estimate.
5. Also compute adjusted R square.
6. Test the significance of each of the regression coefficients.
7. Test the overall goodness of fit of the model.

Solution:



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

From the coefficient table,

1. Y= 6.91 – 0.78x1+0.01x2.
2. 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.

1. R Square = 0.88 i.e. 88% of total deviation on lifetime is explained by two independent variables.
2. Standard error = 0.90 i.e. the average deviation from the fitting regression line is 0.90.
3. Adjusted R square = 0.82
4. Test for B1,

Hypothesis:

H0: The regression coefficient isn’t significant.

H1: The regression coefficient is significant.

Alpha = 5%

Test statistics:

T=2.66

P value =0.056

The 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.

Alpha = 5%

Test statistics:

T=0.04

P value =0.96

The decision, since the p-value is greater than the alpha so we accept H0.

Hence, we conclude that the regression coefficient isn’t significant.

1. Test for regression model

Hypothesis:

H0: The regression model isn’t significant.

H1: The regression model is significant.

Alpha = 5%

Test statistics:

F =15.14

P value =0.01

The decision, since the p-value is smaller than alpha we don’t reject H0.

Hence, we conclude that the regression model isn’t significant.