**Practical No: 02**

**OBJECT:** The following information has been gathered from a random sample of apartment renters in a city. We are trying to predict rent (in dollars per month) based on the size of apartment (number of rooms) and the distance from downtown (in miles).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Rent (Dollar) | 360 | 1000 | 450 | 525 | 350 | 300 |
| Number of rooms | 2 | 6 | 3 | 4 | 2 | 1 |
| Distance from downtown | 1 | 1 | 2 | 3 | 10 |  |

1. Obtain the multiple regression models that best relate these variables
2. Interpret the obtain regression coefficients.
3. If someone is looking for a two bed apartment 2 miles from down town, what rent should he expected to pay?
4. Calculate Total Sum of Square, Sum of Square due to regression and Sum of Square due to Error.
5. Calculate the standard error of estimate.
6. Calculate the Multiple correlation coefficient.
7. Calculate the Coefficient of multiple determination and Adjusted Coefficient of determination. Also interpret your result.
8. Calculate the predicted and residual values.
9. Test the significance of regression coefficient and overall fit of the regression equation.

**WORKING EXPRESSION:**

Let y be the dependent variable and x1 and x2 are the independent variables then the regression equation of y on x1 and x2 is,

Y = a + b1x1 + b2x2

Where, y = dependent variable

x1, x2 = independent variable

b1, b2 = regression coefficient

Total sum of square (T.S.S) =

Sum of square due to error (S.S.E) =

Sum of square due to Regression (S.S.R) = T.S.S - S.S.E

**Standard Error of Estimate (Se):** Standard error is the square root of the variance computed from sample data. The standard error of the estimate measures the average variation or scatteredness of the observed data point around regression line. Standard error of the estimate is used to measure the reliability of the regression equation. Regression line having less standard error of estimate is more reliable than regression line having more standard error of estimate.

Standard Error of Estimate (Se)

Where, SSE = sum of square due to error

k = number of independent variables in regression model

n = number of observations.

**Adjusted R square**: Adjusted is simply a adjusted by its degree of freedom and reflects both the number of independent variables and sample size used in the model. Adjusted is considered as an important measure for the comparising of two or more regression models that predict same dependent variable with different number of independent variables.

**()**

Were,

no. of pair of observations

no. of independent variables.

**Multiple Correlation:** Multiple correlation studies the relationship between dependent variables and joint (or combined) effects of independent variables. For example, multiple correlation gives the relationship of dependent variable (yield of paddy) and joint effect of independent variables (plot of land, labor, seed, fertilizer, pesticide, irrigation and so on.)

**Coefficient of multiple determination:** The square of multiple correlation coefficient is known as coefficient of multiple determination. i.e., , , *.*

Practical No: 02

Name: Aakash Shrestha

Roll No.: 02

Subject: Statistics

Date: 2080/04/07

Faculty: BSc. CSIT 3rd Semester

**OUTPUT:**

REGRESSION

/DESCRIPTIVES MEAN STDDEV CORR SIG N

/MISSING LISTWISE

/STATISTICS COEFF OUTS R ANOVA CHANGE

/CRITERIA=PIN(.05) POUT(.10)

/NOORIGIN

/DEPENDENT Rent

/METHOD=ENTER Number\_of\_rooms Distance\_from\_towndown

/SAVE PRED ADJPRED RESID.

**Regression**

|  |  |  |  |
| --- | --- | --- | --- |
| **Descriptive Statistics** | | | |
|  | Mean | Std. Deviation | N |
| Y | 497.5000 | 258.91601 | 6 |
| X1 | 3.0000 | 1.78885 | 6 |
| X2 | 3.5000 | 3.39116 | 6 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Correlations** | | | | |
|  | | Y | X1 | X2 |
| Pearson Correlation | Y | 1.000 | .956 | -.436 |
| X1 | .956 | 1.000 | -.429 |
| X2 | -.436 | -.429 | 1.000 |
| Sig. (1-tailed) | Y | . | .001 | .194 |
| X1 | .001 | . | .198 |
| X2 | .194 | .198 | . |
| N | Y | 6 | 6 | 6 |
| X1 | 6 | 6 | 6 |
| X2 | 6 | 6 | 6 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | |
|  |  |  |  |
|  |  |  |  |
|  | | | |
|  | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables Entered/Removeda** | | | |
| Model | Variables Entered | Variables Removed | Method |
| 1 | X2, X1b | . | Enter |
| a. Dependent Variable: Y | | | |
| b. All requested variables entered. | | | |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Summaryb** | | | | | | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | | |
| R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .957a | .916 | .859 | 97.08638 | .916 | 16.280 | 2 | 3 | .025 |
| a. Predictors: (Constant), X2, X1 | | | | | | | | | | |
| b. Dependent Variable: Y | | | | | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 306910.203 | 2 | 153455.102 | 16.280 | .025b |
| Residual | 28277.297 | 3 | 9425.766 |  |  |
| Total | 335187.500 | 5 |  |  |  |
| a. Dependent Variable: Y | | | | | | |
| b. Predictors: (Constant), X2, X1 | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | 96.458 | 118.121 |  | .817 | .474 |
| X1 | 136.485 | 26.864 | .943 | 5.081 | .015 |
| X2 | -2.403 | 14.171 | -.031 | -.170 | .876 |
| a. Dependent Variable: Y | | | | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Residuals Statisticsa** | | | | | |
|  | Minimum | Maximum | Mean | Std. Deviation | N |
| Predicted Value | 223.3289 | 912.9627 | 497.5000 | 247.75399 | 6 |
| Std. Predicted Value | -1.107 | 1.677 | .000 | 1.000 | 6 |
| Standard Error of Predicted Value | 44.975 | 93.037 | 66.388 | 19.146 | 6 |
| Adjusted Predicted Value | 162.1939 | 678.5328 | 448.4655 | 208.48503 | 6 |
| Residual | -110.18642 | 87.03728 | .00000 | 75.20279 | 6 |
| Std. Residual | -1.135 | .896 | .000 | .775 | 6 |
| Stud. Residual | -1.294 | 1.723 | .160 | 1.096 | 6 |
| Deleted Residual | -143.33141 | 321.46722 | 49.03454 | 164.94651 | 6 |
| Stud. Deleted Residual | -1.591 | 13.711 | 2.125 | 5.743 | 6 |
| Mahal. Distance | .240 | 3.758 | 1.667 | 1.389 | 6 |
| Cook's Distance | .003 | 2.665 | .545 | 1.044 | 6 |
| Centered Leverage Value | .048 | .752 | .333 | .278 | 6 |
| a. Dependent Variable: Y | | | | | |

**RESULTS:**

1. The multiple regression model of y on x1 and x2 is,

Y = 96.458 + 136.484X1 - 2.403X2

1. Interpretation:

b1 = 136.484 means on average rent is increased by 136.484 when room is increased by 1 holding me effect of distance from downtown constant.

b2 = -2.403 means average rent is decreased by 2.403 when the distance home downtown is increased by 1 holding the effect of number of room constant.

1. When x1 = 2 and x2 = 2

Y = 96.458 + 136.484X1 - 2.403X2

= 96.458 +136.484\*2 - 2.403 \*2

= 364.62

Therefore, expected rent for two-bedroom apartment 2 miles from downtown is 364.62 dollar.

1. Total sum of Square (T.S.S) = = 335187.500

Sum of square due to error (S.S.E) = = 28277.297

Sum of square due to regression (S.S.R) = 306910.203

1. Standard error of estimate (Se) = = 97.08638
2. Multiple correlation coefficient (R) = 0.957
3. Coefficient of multiple determination (R²) =0.916

Interpretation: It means 91.6% of total variation in rent (y) can be explained by the variation in rooms and distance from downtown.

Adjusted R square: 0.859

Interpretation: This indicates this regression equation can represent 85.9% of the variation in the dependent variable (response variable) is explained by the independent variables.

VIII. The predicted and residual value are:

|  |  |  |
| --- | --- | --- |
| S. N | Predicted | Residual |
| 1 | 367.02937 | -7.02397 |
| 2 | 912.96272 | 87.03728 |
| 3 | 501.10519 | -51.10519 |
| 4 | 635.18642 | -110.18642 |
| 5 | 345.3928 | 4.60719 |
| 6 | 223.3300 | 76.6700 |

IX.  As from ANOVA Table we can clearly see the value of Fcal = 16.280 also p-value = 0.005.

Since, level of significance is 5% = 0.05.

**Decision:** p-value = 0.025 < 0.05

Hence, we reject Ho. Which means there is no significant linear relationship between regression equation.

**CONCLUSION:**

Hence, we have obtained the regression model equation and interpret it. We have calculated sum of square due to error, Total sum of square, sum of square due to regression, standard error of estimate. We have calculated and interpreted the value of coefficient of multiple determination and Adjusted R2. We have calculated the predicted value and residual value from variable view and at last we have test the significance of regression equation.