Lab no. 1:

Independent t-test:

The score obtained by the two group of the students is given in the table:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Group A | 44 | 30 | 35 | 20 | 28 | 25 |
| Group B | 22 | 29 | 38 | 24 | 27 | 32 |

At 5% level of the significance, test whether there is significance difference between the average score of two group of students.

Solution,

Hypothesis:

H0 : µ1 = µ2

H1 : µ1 ≠ µ2

Level of significance:

Apha = 5%

Test statistics:

|  |  |  |
| --- | --- | --- |
| t-Test: Two-Sample Assuming Equal Variances |  |  |
|  |  |  |
|  | *Variable 1* | *Variable 2* |
| Mean | 30.33333333 | 28.66666667 |
| Variance | 69.86666667 | 33.46666667 |
| Observations | 6 | 6 |
| Pooled Variance | 51.66666667 |  |
| Hypothesized Mean Difference | 0 |  |
| df | 10 |  |
| t Stat | 0.401609664 |  |
| P(T<=t) one-tail | 0.348209933 |  |
| t Critical one-tail | 1.812461123 |  |
| P(T<=t) two-tail | 0.696419867 |  |
| t Critical two-tail | 2.228138852 |  |

Decision:

Since, t(cal) < t(tab) so we donot reject H0.

Hence, we conclude that there is no significance difference between the average score of two group of students.

Lab no. 2:

The problem solving time of theto group of student is given below:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| A | 16 | 9 | 12 | 18 | 11 |  |  |
| B | 17 | 18 | 10 | 12 | 14 | 20 | 13 |

At 5% of the level of significance, test problem solving time of the group B is more than the Group A.

Solution,

Hypothesis:

H0 : µ1 = µ2 i.e. the problem solving time of Group B is not higher than the Gorup A

H1 : µ1 > µ2 i.e. the problem solving time of Group B is higher than the Gorup A

Level of the significance:

Alpha = 5%

Test statistics:

|  |  |  |
| --- | --- | --- |
| t-Test: Two-Sample Assuming Equal Variances |  |  |
|  |  |  |
|  | *Variable 1* | *Variable 2* |
| Mean | 13.2 | 14.85714286 |
| Variance | 13.7 | 12.80952381 |
| Observations | 5 | 7 |
| Pooled Variance | 13.16571429 |  |
| Hypothesized Mean Difference | 0 |  |
| df | 10 |  |
| t Stat | -0.77997581 |  |
| P(T<=t) one-tail | 0.226735184 |  |
| t Critical one-tail | 1.812461123 |  |
| P(T<=t) two-tail | 0.453470368 |  |
| t Critical two-tail | 2.228138852 |  |

Decision:

Since, t(cal) < t(tab) so we donot reject H0.

Hence, the problem solving time of Group B is not higher than the Gorup A.

Lab no . 3:

Paired t-test:

The performance score the students before and after training is given below.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Before | 14 | 22 | 30 | 18 | 24 | 28 | 20 |
| After | 16 | 21 | 34 | 13 | 22 | 32 | 23 |

At alpha equals to 5%, test whether the training is beneficial or not.

Solution:

Hypothesis:

H0 : µ1 = µ2 i.e. the training is not beneficial.

H1 : µ1 < µ2 i.e. the training is beneficial.

Level of the significance:

Alpha = 5%

Test Statistics:

|  |  |  |
| --- | --- | --- |
| t-Test: Paired Two Sample for Means |  |  |
|  |  |  |
|  | *Variable 1* | *Variable 2* |
| Mean | 22.28571429 | 23 |
| Variance | 31.23809524 | 59.33333333 |
| Observations | 7 | 7 |
| Pearson Correlation | 0.913627381 |  |
| Hypothesized Mean Difference | 0 |  |
| df | 6 |  |
| t Stat | -0.54772256 |  |
| P(T<=t) one-tail | 0.301822528 |  |
| t Critical one-tail | 1.943180281 |  |
| P(T<=t) two-tail | 0.603645057 |  |
| t Critical two-tail | 2.446911851 |  |

Decision:

Since, t(cal) < t(tab) so we donot reject H0.

Hence, the training is not benifical.

Lab 4:

The problem solving time of the group of student is given below:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Before | 16 | 20 | 21 | 18 | 17 | 22 | 30 |
| After | 16 | 24 | 20 | 20 | 18 | 24 | 29 |

At 5% of the level of significance, test problem solving time of the group B is more than the Group A.

Solution,

Hypothesis:

H0 : µ1 = µ2 i.e. the problem solving time of Group B is not higher than the Gorup A

H1 : µ1 > µ2 i.e. the problem solving time of Group B is higher than the Gorup A

Level of the significance:

Alpha = 5%

Test statistics:

|  |  |  |
| --- | --- | --- |
| t-Test: Paired Two Sample for Means |  |  |
|  |  |  |
|  | *Variable 1* | *Variable 2* |
| Mean | 20.57142857 | 21.57142857 |
| Variance | 21.95238095 | 19.28571429 |
| Observations | 7 | 7 |
| Pearson Correlation | 0.921096427 |  |
| Hypothesized Mean Difference | 0 |  |
| Df | 6 |  |
| t Stat | -1.44913767 |  |
| P(T<=t) one-tail | 0.098736513 |  |
| t Critical one-tail | 1.943180281 |  |
| P(T<=t) two-tail | 0.197473026 |  |
| t Critical two-tail | 2.446911851 |  |

Decision:

Since, t(cal) < t(tab) so we donot reject H0.

Hence, the training is not benifical.

Lab 5:

An analyst predicting user engagement (measured as time spent in minutes per day) in an online games based on two factors : the number of friends in the game the level of number of friends in the game, the level of the user (1 to 50), the dataset is as follows:

|  |  |  |
| --- | --- | --- |
| Number of friends | Level | Engagement(minutes) |
| 15 | 5 | 45 |
| 19 | 9 | 60 |
| 28 | 16 | 65 |
| 30 | 26 | 85 |
| 37 | 30 | 90 |

1. Compute the multiple correlation coefficient of engagement with number of friend and the level of the game.
2. Fit a multiple regression model to describe the given variables
3. Interpret the meaning of estimated regression coefficient
4. Compute TSS, SSR and SSE.
5. Compute the coefficient of the determination and standard error of the estimate with interpretation.
6. Test the significance of coefficient of number of friends.
7. Test the overall goodness of fit of the regression model.

Solution,

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT | |  |  |  |  |
|  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |
| Multiple R | 0.986179993 |  |  |  |  |
| R Square | 0.972550978 |  |  |  |  |
| Adjusted R Square | 0.945101956 |  |  |  |  |
| Standard Error | 4.33619421 |  |  |  |  |
| Observations | 5 |  |  |  |  |
|  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 2 | 1332.39484 | 666.1974198 | 35.43117 | 0.027449 |
| Residual | 2 | 37.60516045 | 18.80258023 |  |  |
| Total | 4 | 1370 |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficients | Standard Error | t Stat | P-value |
| Intercept | 44.33933218 | 12.36996301 | 3.584435308 | 0.069783 |
| X Variable 1 | -0.381396357 | 0.967076394 | -0.3943808 | 0.73138 |
| X Variable 2 | 2.005854293 | 0.795956957 | 2.520053723 | 0.127934 |

1. the multiple correlation coefficient Of engagement with number of friend and the level of the game is 0.98.
2. let the multiple regression model be,

y = a + b1\*x1 + b2\*x2

from coefficient table, fitted regression equation be,

y = 44.33 - 0.38\*x1 + 2.00\*x2

1. b1 = -0.38 means, when the number of friends increases by 1 unit then the engagement time decreases by 0.38 unit, keeping the effect of level of the game as constant.

b2 = 2.00 means, when the level of the game increases by 1 unit then the engagement time increases by 2.00 unit, keeping the effect of number of friends in the game as constant.

1. From the ANOVA table,

TSS = 1370

SSR = 1332.39

SSE = 37.60

1. the coefficient of the determination r2 = 0.9725, that is, 97.25% of the total variation on the engagement time is explained by two independent variables number of friends and level of the game.

The standard error of the estimate with interpretation is 4.33, i.e. the average deviation of the observation from the fitted regression line is 4.33.

1. test for b1,

Hypothesis:

H0 : regression coefficient is not significant

H1 : regression coefficient is significant

Level of significance:

Alpha = 5%

Test statistics:

|t| = b1/sb1 = 0.39

Probability value:

P = 0.73

Decision:

Since, p(value) > alpha so we donot reject H0.

Hence we conclude that the regression coefficient is not significant

1. Test the overall goodness of fit of the regression model.

Hypothesis:

H0 : regression model is not significant

H1 : regression model is significant

Level of significance:

Alpha = 5%

Test statistics:  
f = 35.43

Probability value:

P = 0.027

Decision:

Since, p(value) < alpha so we reject H0.

Hence we conclude that the regression model is significant.

Lab 6:

A software developer team wants to predict the number of bugs(Y) in a software system based on two factors: number of developers (X1), and number of testing hours(X2). The dataset is a follows:

|  |  |  |
| --- | --- | --- |
| Number of developers | Testing hours | Bug Count |
| 5 | 30 | 10 |
| 9 | 36 | 12 |
| 11 | 45 | 15 |
| 16 | 60 | 21 |
| 19 | 65 | 25 |

1. Compute the multiple correlation coefficient of number of bugs with number of developer and the testing hours
2. Fit a multiple regression model to describe the given variables
3. Interpret the meaning of estimated regression coefficient
4. Compute TSS, SSR and SSE.
5. Compute the coefficient of the determination and standard error of the estimate with interpretation.
6. Test the significance of coefficient of testing hours.
7. Test the overall goodness of fit of the regression model.

Solution,

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT | |  |  |  |  |
|  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |
| Multiple R | 0.992757085 |  |  |  |  |
| R Square | 0.98556663 |  |  |  |  |
| Adjusted R Square | 0.97113326 |  |  |  |  |
| Standard Error | 1.06511167 |  |  |  |  |
| Observations | 5 |  |  |  |  |
|  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 2 | 154.9310743 | 77.46553713 | 68.28389 | 0.01443337 |
| Residual | 2 | 2.268925739 | 1.13446287 |  |  |
| Total | 4 | 157.2 |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
| Intercept | -0.894736842 | 3.992767054 | -0.22408942 | 0.843497 |
| X Variable 1 | 0.377433309 | 0.680851936 | 0.554354463 | 0.635049 |
| X Variable 2 | 0.274693583 | 0.251772194 | 1.091040195 | 0.38917 |

1. the multiple correlation coefficient of bug count with number of developer and the testing hours is 0.99.
2. let the multiple regression model be,

y = a + b1\*x1 + b2\*x2

from coefficient table, fitted regression equation be,

y = -0.89 + 0.37\*x1 + 0.27\*x2

1. b1 = 0.37 means, when the number of developers increases by 1 unit then the bug count decreases by 0.38 unit, keeping the effect of testing hour as constant.

B2 = 0.27 means, when the testing hour increases by 1 unit then the bug count decreases by 0.27 unit, keeping the effect of bug count as constant.

1. From the ANOVA table,

TSS = 157.2

SSR = 154.93

SSE = 2.26

1. the coefficient of the determination r2 = 0.9725, that is, 97.25% of the total variation on the engagement time is explained by two independent variables number of friends and level of the game.

The standard error of the estimate with interpretation is 4.33, i.e. the average deviation of the observation from the fitted regression line is 4.33.

1. test for b1,

Hypothesis:

H0 : regression coefficient is not significant

H1 : regression coefficient is significant

Level of significance:

Alpha = 5%

Test statistics:

|t| = b1/sb1 = 0.39

Probability value:

P = 0.73

Decision:

Since, p(value) > alpha so we donot reject H0.

Hence we conclude that the regression coefficient is not significant

1. Test the overall goodness of fit of the regression model.

Hypothesis:

H0 : regression model is not significant

H1 : regression model is significant

Level of significance:

Alpha = 5%

Test statistics:  
f = 35.43

Probability value:

P = 0.027

Decision:

Since, p(value) < alpha so we reject H0.

Hence we conclude that the regression model is significant.

Lab 7:

The following data set represents the layout of CRD of three treatments each replicated 4 times.

|  |  |  |  |
| --- | --- | --- | --- |
| A9 | B11 | C8 | C7 |
| B16 | A14 | B12 | B14 |
| C12 | C16 | A13 | A6 |

Carry out analysis of the design.

Solution,

Hypothesis:

H0 : there is no significance between the treatments

H1 : there is significance between the treatments

Level of significance:

Alpha = 5%

Test statistics:

Treatment

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Anova: Single Factor |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Row 1 | 4 | 42 | 10.5 | 13.66667 |  |  |
| Row 2 | 4 | 53 | 13.25 | 4.916667 |  |  |
| Row 3 | 4 | 43 | 10.75 | 16.91667 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Treatment | 18.5 | 2 | 9.25 | 0.78169 | 0.486382764 | 4.256495 |
| Error | 106.5 | 9 | 11.83333333 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 125 | 11 |  |  |  |  |

Decision:

Since, f(cal) < f(tab), so we don’t reject H0.

Hence, there is no significance between the treatments.

Lab 8:

What do you understand by “Design of an Experiment”? Physicians depend the laboratory test results when managing the medical problems such as diabetes or epilepsy. In an uniformity test glucose tolerance, three different laboratories were sent nt=5 identical blood samples from a person who had drunk 50 mg. of glucose dissolved in water. The laboratory results are listed below:

|  |  |  |
| --- | --- | --- |
| **Lab 1** | **Lab 2** | **Lab 3** |
| 12.1 | 9.3 | 10.0 |
| 11.7 | 11.1 | 10.5 |
| 10.9 | 10.7 | 10.1 |
| 10.2 | 10.9 | 11.0 |
| 10.6 | 9.0 | 10.4 |

Do data indicate a difference in the average readings for three laboratories? Use α=.05

Solution,

Here

Hypothesis:

H0 : there is no significance difference between the average reading of the laboratory

H1 : there is significance difference between the average reading of the laboratory

Level of significance :

Alpha = 5%

Test statistics:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Anova: Single Factor |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Column 1 | 5 | 55.5 | 11.1 | 0.615 |  |  |
| Column 2 | 5 | 51 | 10.2 | 0.95 |  |  |
| Column 3 | 5 | 52 | 10.4 | 0.155 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 2.233333333 | 2 | 1.116666667 | 1.947674 | 0.185125863 | 3.885294 |
| Within Groups | 6.88 | 12 | 0.573333333 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 9.113333333 | 14 |  |  |  |  |

Decision:

Since, f(cal) < f(tab), so we don’t reject H0.

Hence, there is no significance difference between the average reading of the laboratory.