**AD8412 DATA ANALYTICS LABORATORY RECORD**

**AD8412 DATA ANALYTICS LABORATORY**

## **INDEX**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Expt.  No. | Name of the Experiment | Page  No | Date | Signature |
| 1 | Demonstration the random sampling |  |  |  |
| 2 | Demonstrate the probability sampling from a known population |  |  |  |
| 3 | Implementation of Z-Test – One Sample Z-Test and Two Sample Z-Test |  |  |  |
| 4 | Implementation of Z-Test – using Titanic case study |  |  |  |
| 5 | Implementation of T-Test – one sample t-test |  |  |  |
| 6 | Implementation of T-Test – Two sample t-test and  Paired T-Test |  |  |  |
| 7 | Implementation of VARIANCE ANALYSIS ( ANNOVA) |  |  |  |
| 8 | Demonstration of Linear Regression |  |  |  |
| 9 | Demonstration of Logistic Regression |  |  |  |
| 10 | Demonstration of Multiple-Linear Regression |  |  |  |
| 11 | Implementation of Time Series Analysis |  |  |  |

|  |  |
| --- | --- |
| **Ex.No: 1**  **Date :** | **Demonstration the random sampling using python** |

**AIM:**

Write a python program to Demonstration the random sampling.

**ALOGRITHM:**

1. Create a list
2. Use the ‘random.choice()’ method to randomly select individual values from this list
3. Use ‘random.sample()’ method for randomly sampling N items from a list
4. In addition to random selection and sampling, the random module has a function for shuffling items in a list
5. Picking Random Items in a List using ‘random.sample()’
6. The random module has a function for generating a random integer provided a range of values. Perform Randomly Shuffling Items in a List using ‘random.shuffle()
7. The random module has a function for generating a random integer provided a range of values. Generatethe Random Integers using ‘random.randint()’
8. The random module also has a function for generating a random floating point value between 0 and 1. Generate the Random Floating Point Values
9. Scale the random float numbers. If we want random numbers between 0 and 500 we just multiply our random number by 500
10. And if we want to add a lower bound as well we can add a conditional statement before appending
11. The random module has a function for computing uniformly distributed numbers. Compute Uniformly Distributed Numbers with ‘random.uniform()’
12. The random module has a function for computing normally distributed numbers. Compute Normally Distributed Numbers with ‘random.gauss()’

**RESULT:**

Thus the python program to perform random sampling has been implemented and executed sucessfully.

|  |  |
| --- | --- |
| **Ex.No: 2**  **Date :** | **Demonstration the probability sampling using python** |

**AIM :**

Write a python program to Demonstration the probability sampling.

**ALGORITHM:**

* 1. Create a sample from a set of 10 products using probability sampling to determine the population mean of a particular measure of interest.
  2. Implement Simple Random Sampling

The simple random sampling method selects random samples from a process or population where every unit has the same probability of getting selected

* 1. Implement Systematic Sampling

The systematic sampling method selects units based on a fixed sampling interval

* 1. Implement Cluster Sampling

The cluster sampling method divides the population in clusters of equal size n and selects clusters every Tth time

* 1. Implement Stratified Random Sampling

The stratified random sampling method divides the population in subgroupsand selects random samples where every unit has the same probability of getting selected

**RESULT:**

Thus the python program to perform probability sampling has been implemented and executed succesfully.

|  |  |
| --- | --- |
| **Ex.No: 3**  **Date :** | **Implementation of Z-Test – One Sample Z-Test and Two Sample Z-Test** |

**AIM:**

Write a python program to Implementation of Z-Test – One Sample Z-Test and Two Sample Z-Test.

**ALGORITHM:**

1. Evaluate the data distribution.
2. Formulate Hypothesis statement symbolically
3. Define the level of significance (alpha)
4. Calculate Z test statistic or Z score.
5. Derive P-value for the Z score calculated.
6. Make decision:

6.1. P-Value <= alpha, then we reject H0.

6.2. If P-Value > alpha, Fail to reject H0.

**RESULT:**

Thus the python program to perform Z-Test – One Sample Z-Test and Two Sample Z-Test has been implemented and executed successfully

|  |  |
| --- | --- |
| **Ex.No: 4**  **Date :** | **Implementation of Z-Test – using Titanic case study** |

**AIM:**

Write a python program to Implementation of Z-Test – using Titanic case study.

**ALGORITHM:**

1. Evaluate the data distribution.
2. Formulate Hypothesis statement symbolically
3. Define the level of significance (alpha)
4. Calculate Z test statistic or Z score.
5. Derive P-value for the Z score calculated.
6. Make decision:

6.1. P-Value <= alpha, then we reject H0.

6.2. If P-Value > alpha, Fail to reject H0.

**RESULTS:**

Thus the python program to perform Z-Test – using Titanic case study has been implemented and excecuted successfully.

|  |  |
| --- | --- |
| **Ex.No: 5**  **Date :** | **Implementation of T-Test – one sample t-test** |

**AIM:**

Write a python program Implementation of T-Test – one sample t-test.

**ALGORITHM:**

1. Create some dummy age data for the population of voters in the entire country
2. Create Sample of voters in Minnesota and test the whether the average age of voters Minnesota differs from the population
3. Conduct a t-test at a 95% confidence level and see if it correctly rejects the null hypothesis that the sample comes from the same distribution as the population.
4. If the t-statistic lies outside the quantiles of the t-distribution corresponding to our confidence level and degrees of freedom, we reject the null hypothesis.
5. Calculate the chances of seeing a result as extreme as the one being observed (known as the p-value) by passing the t-statistic in as the quantile to the stats.t.cdf() function

**RESULT :**

Thus the python program to perform T-Test – one sample t-test has been implemented and executed successfully

|  |  |
| --- | --- |
| **Ex.No: 6**  **Date :** | **Implementation of T-Test – Two sample t-test and**  **Paired T-Test** |

**AIM:**

Write a python program to implementation of T-test – Two sample t-test and Paried T-test.

**ALGORITHM:**

1 Create the data

2 Conduct a two sample t-test.

1. Interpret the results

**RESULT :**

Thus the python program to perform T-Test – Two sample t-test andPaired T-Test has been implemented and executed successfully

|  |  |
| --- | --- |
| **Ex.No: 7**  **Date :** | **Implementation of Variance Analysis ( ANOVA)** |

**AIM :**

Write a python program to implementation of variance analysis(ANOVA).

**ALGORITHM:**

A. Input:

A bunch of students from different colleges taking the same exam. You want to see if one college outperforms the other, hence your null hypothesis is that the meansof GPAs in each group are equivalent to those of the other groups. To keep it simple,we will consider 3 groups (college ‘A’, ‘B’, ‘C’) with 6 students each.

A=[25,25,27,30,23,20]

B=[30,30,21,24,26,28]

C=[18,30,29,29,24,26]

Null Hypothesis: GPAs in each group are equivalent to those of the other groups.

Alternate Hypothesis – There is a significant difference among the groups

B. Output:

To find the null hypothesis or alternate hypothesis is acceptable or not.

1. Rows are grouped according to their value in the category column.

2. The total mean value of the value column is computed.

3. The mean within each group is computed.

4. The difference between each value and the mean value for the group is calculated and squared.

5. The squared difference values are added. The result is a value that relates tothe total deviation of rows from the mean of their respective groups. This valueis referred to as the sum of squares within groups, or S2Wthn.

6. For each group, the difference between the total mean and the group mean issquared and multiplied by the number of values in the group. The results areadded. The result is referred to as the sum of squares between groups or S2Btwn.

7. The two sums of squares are used to obtain a statistic for testing the nullhypothesis, the so called F-statistic. The F-statistic is calculated as:

wheredfBtwn (degree of freedom between groups) equals the number of groupsminus 1, and dfWthn (degree of freedom within groups) equals the totalnumber of values minus the number of groups

8. The F-statistic is distributed according to the F-distribution (commonlypresented in mathematical tables/handbooks). The F-statistic, in combinationwith the degrees of freedom and an F-distribution table, yields the p-value.

The p-value is the probability of the actual or a more extreme outcome underthe null-hypothesis. The lower the p-value, the larger the difference

**RESULT:**

Thus the python program to perform ANOVA has been implemented and executed successfully

|  |  |
| --- | --- |
| **Ex.No: 8**  **Date :** | **Demonstration of Linear Regression** |

**AIM:**

Write a python program to demonstration of linear regression.

**ALGORITHM:**

A. Input: Get any value of x.

B.Output: Find the value of y for any x.

1. Consider a set of values x, y.

2 Take the linear set of equation y = a+bx.

3 Computer value of a, b with respect to the given values, b = nΣxy − (Σx)

(Σy) / nΣx2−(Σx)2, a = Σy−b (Σx)n.

4. Implement the value of a, b in the equation y = a+ bx.

5. Regress the value of y for any x.

**RESULT:**

Thus the python program to perform linear regression has been implemented and executed successfully

|  |  |
| --- | --- |
| **Ex.No: 9**  **Date :** | **Demonstration of Logistic Regression** |

**AIM:**

Write a python program to demonstration of logistic regression.

**ALGORITHM:**

A. Input: GMAT score, GPA and Years of work experience directly given in the program as input.

B. Output: Aspiring candidate get admitted or not.

1.Initialize the variables

2. Set the Data frame

3. Spilt data set into training and testing.

4.Fit the data into logistic regression function.

5. Predict the test data set.

6. Print the results.

**RESULT:**

Thus the python program to perform logistic regression has been implemented and executed successfully.

.

|  |  |
| --- | --- |
| **Ex.No: 10**  **Date :** | **Demonstration of Multiple-Linear Regression** |

**AIM :**

Write a python program to demonstration of multiple-linear regression.

**ALGORITHM:**

A. Input: Boston house pricing dataset using Scikit-learn.

B. Output: List of all Coefficients, variance score and residual error plots

1. Get the multi-attribute dataset using the Scikit-learn data source.
2. Create a regression object.
3. Train the dataset with the regression model fit.
4. Get and print the regression coefficients and variance.

5. Plot the residual error.

**RESULT:**

Thus the python program to perform multiple-linear regression has been implemented and executed successfully

|  |  |
| --- | --- |
| **Ex.No: 11**  **Date :** | **Implementation of Time Series Analysis** |

**AIM :**

Write a python program to implementation of time series analysis.

**ALGORITHM:**

1. Loading time series dataset correctly in Pandas

2. Indexing in Time-Series Data

3. Time-Resampling using Pandas

4 Rolling Time Series

5 Plotting Time-series Data using Pandas

**RESULT:**

Thus the python program to perform time series analysis has been implemented and executed successfully.