## **PROJECT 3: TOPIC 5**

# **Report**

* IRIS.xlsx file was converted to IRIS.csv to read the data easily.

**File:** Project3.R

Data from ‘IRIS.csv’ is read into ‘data’.

Since there are 3 species, the data is further categorized into 3 dataframes, namely:

* ‘setosa.data’ – species: Iris-satosa
* ‘versicolor.data’ – species: Iris-versicolor
* ‘virginica.data’ – species: Iris-virginica

Now using ANOVA we get following F-values for respective attributes:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Attribute** | SepalLengthCm | SepalWidthCm | PetalLengthCm | PetalWidthCm |
| **F-value** | 119.2645 | 47.36446 | 1179.034 | 959.3244 |

And the critical values for F-distribution are:

* With confidence level 1% - 99.49236
* With confidence level 5% - 19.48892

Hypothesis H0: Mean values of each species for a given attribute are same/equal. So, the above hypothesis is false for F-value greater than the critical value with given confidence level i.e. mean values are significantly different.

Following is the result if means are significantly different or not:

|  |  |  |
| --- | --- | --- |
| **Attributes** | **Confidence Level 1%** | **Confidence Level 5%** |
| SepalLengthCm | Yes | Yes |
| SepalWidthCm | No | Yes |
| PetalLengthCm | Yes | Yes |
| PetalWidthCm | Yes | Yes |