Probability Project

# Data Analysis

**Group Members:**

Rahim Muhammad Syed 18K-0122 CS-2H

Syed Abdullah Muzaffar 18K-0169 CS-2H

Muhammad Ahmed Khan 18K-1103 CS-2H

**Introduction:**

The Project consists of the following dataset:

“iris”

The aforementioned dataset can be found in the base R package.

The data consists of 5 Columns:

"Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width" "Species"

where Species is categorical, and the rest are numerical observations.

Analysis of the dataset was performed using R Language.

**Methods:**

Hypothesis Testing:

Null Hypothesis: The mean “Sepal Length” of the two species:

* 1. “Versicolor”
  2. “Setosa”

should be equal.

The t.test(data1,data2) function was used to calculate the p.value for the Hypothesis test.

ANOVA:

Linear Regression:

We assumed y=Sepal.Length, and x1…xn as the remaining numerical observations, to find the Linear Regression Line for “Sepal.Length” against "Sepal.Width", "Petal.Length", "Petal.Width" .

The lm(formula=y~x1+x2.. , dataset) function was used to calculate the linear regression model.

**Result:**

Hypothesis Testing:

ANOVA:

Linear Regression:

The R script file for the methods and results shown is included in the folder.