****

**ABHINAV VIJAYAKUMAR**

**19BCE1311**

**CSE3506 – ESSENTIALS OF DATA ANALYTICS LAB-4**

**DR. LAKSHMI PATHI JAKKAMPUTI (L21 + L22)**

**--------------------------------------------------------------------------------------------**

**Tasks for Week-4: Analysis of Variance (ANOVA)**

**Perform ANOVA test and determine the statistical differences between the means of individual groups given in the data**

**Aim**: Perform ANOVA test and determine the statistical differences between the means of individual groups given in the data.

**Algorithm:**

**1.** Import the dataset and load the dplyr library.

**2.** Group the data using the group\_by function based on color.

**3.** Apply ANOVA using response with respect to color and generate summary.

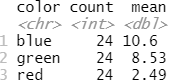
**4.** If Pr(>F) value< 0.05, then perform the Tukey HSD test.

**5.** If the p-adjusted value of the pair is less than 0.05 then they are significantly different else they are not.

**Statistics:**

**1. Applying group by:**

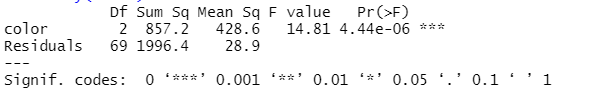
group\_by(data,color) %>% summarise(count = n(),mean = mean(response, na.rm = TRUE))



**2. Summary of ANOVA:**

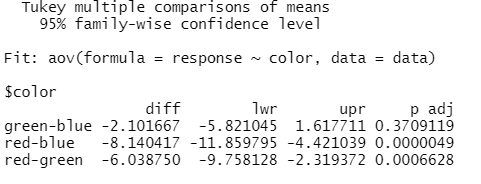
ANOVA <- aov(response~color, data = data)

summary(ANOVA)



**3. Conducting Tukey HSD Test**

TukeyHSD(ANOVA)



**Inference:**

**1.** As seen in the summary of ANOVA, the profit value (Pr(>F)) is less than 0.05, hence the null hypothesis is rejected and the Tukey HSD test is required.

**2.** As seen in the Tukey HSD test results,

**i)** green and blue are not significantly different since p adj is more than 0.05.

**ii)** red and blue are significantly different since p adj is less than 0.05.

**iii)** green and red are significantly different since p adj is less than 0.05.

**Program:**

# To clear the environment

rm(list=ls())

setwd("C:/Users/Abhinav Vijayakumar/Desktop/VIT Academics/Sem 6/EDA/LAB/LAB 4")

data <- read.csv("color-anova-example.csv")

library(dplyr) # To group the data

group\_by(data,color) %>% summarise(count = n(),mean = mean(response, na.rm = TRUE))

# ANOVA

ANOVA <- aov(response~color, data = data)

summary(ANOVA)

TukeyHSD(ANOVA)