**NAME: SANDEEP PATIL**

**Email: saashasandeep@gmail.com**

**DATA ANALYTICS WITH R, EXCEL and TABLEAU**

**Session 12 – Assignment – 12.1**

#1. Use the given link Data Set.

yeastdata <- read.table("D:/BIG DATA/DATA ANALYTICS WITH R, EXCEL & TABLEAU/12 NON-LINEAR MODELS/yeast\_data.txt", quote="\"", comment.char="")

View(yeastdata)

dim(yeastdata)

library(reshape)

yeastdata <- rename (yeastdata, c(V1 = "Sequence Name",

V2 = "mcg",

V3 = "gvh",

V4 = "alm",

V5 = "mit",

V6 = "erl",

V7 = "pox",

V8 = "vac",

V9 = "nuc",

V10 = "Class Distribution"))

View(yeastdata)

#Answer the below questions:

#a. Perform ANOVA test on the discriminant analysis scores of nuclear localization signals of both nuclear and non-nuclear proteins by class variables (Target).

results <- aov (yeastdata$nuc ~ yeastdata$`Class Distribution`)

summary(results)

#Since p value is significantly lesser than 0.05 we have strong evidence to reject null hypothesis

#and conclude that there is at least one class significantly different than others.

#b. Which class is significantly different from others?

plot(yeastdata$nuc ~ yeastdata$`Class Distribution`, col = heat.colors(10), xlab = "Class Distribution", ylab = "Score of discriminant analysis" , main = "Boxplot across classes")

#Nuclear class is significantly different from others.