#----------------------- Assignment 6.3 ----------------------

# Import Data Set

yeast <- read.table("E:/Data Analytics with RET/Assignment/yeast.txt", quote="\"", comment.char="")

View(yeast)

names(yeast) <- c("seq","mcg", "gvh", "alm", "mit", "erl", "pox", "vac", "nuc", "class")

head(yeast)

dim(yeast)

sapply(yeast, function(x) sum(is.na(yeast)))

#---------------------------------------------

# a. Perform ANOVA test on the discriminant analysis scores of nuclear

# localization signals of both nuclear

# and non-nuclear proteins by class variables (Target).

# HO : scores across the class variables is same

x <- aov(nuc~class, data = yeast)

summary(x)

# since p value is less than 0.05, scores across the class variables is

# not same at 95% confidence level

# --------------------------------------------

# b. Which class is significantly different from others?

TukeyHSD(x)

# ERL is significantly different from others

#----------------------------------------------