1. Use the given link Data Set.

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).

ANS.

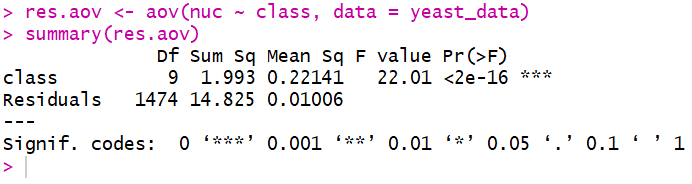
For anova test lets define null and alternative hypothesis.

: The means of nuclear localization signals of the different groups are the same.

: At least one sample mean is not equal to the others.

res.aov <- aov(nuc ~ class, data=yeast\_data)

summary(res.aov)



As the p-value is less than the significance level 0.05, we can conclude that there are significant differences between the groups highlighted with “\*" in the model summary.

b. Which class is significantly different from others?

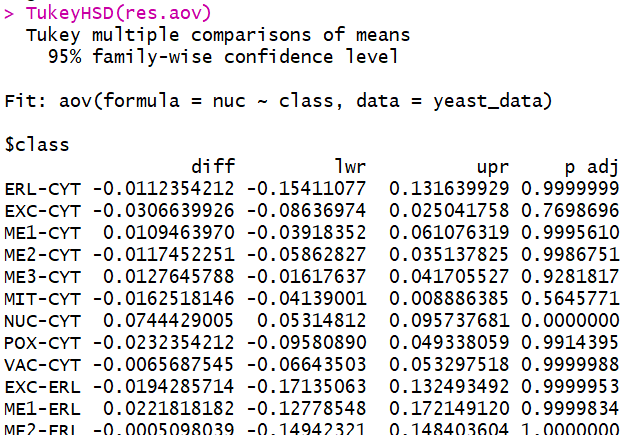
ANS.

In one-way ANOVA test, a significant p-value indicates that some of the group means are different, but we don’t know which pairs of groups are different. It’s possible to perform multiple pairwise-comparison, to determine if the mean difference between specific pairs of group are statistically significant.

As the ANOVA test is significant, we can compute **Tukey HSD** (Tukey Honest Significant Differences, R function: **TukeyHSD**()) for performing multiple pairwise-comparison between the means of groups.

The function **TukeyHSD**() takes the fitted ANOVA as an argument.

TukeyHSD(res.aov)



[ Very large output so CROPPED]

From above output we got following groups for which there is a significant difference between the means.

**NUC-CYT 0.0744429005 0.05314812 0.095737681 0.0000000**

**NUC-EXC 0.1051068931 0.04924622 0.160967571 0.0000001**

**NUC-ME1 0.0634965035 0.01319448 0.113798528 0.0026743**

**NUC-ME2 0.0861881256 0.03912110 0.133255152 0.0000004**

**NUC-ME3 0.0616783217 0.03244028 0.090916364 0.0000000**

**NUC-MIT 0.0906947151 0.06521504 0.116174388 0.0000000**

**POX-NUC -0.0976783217 -0.17037079 -0.024985855 0.0009144**

**VAC-NUC -0.0810116550 -0.14102211 -0.021001196 0.0008397**