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

```
E:/Local Disk/Data Science Training/assign/R-prog/
> col_names <- c("Sequence Name","mcg","gvh","alm","mit","erl","pox","vac","nuc","class")
> datr=read.table(url("https://archive.ics.uci.edu/ml/machine-learning-databases/yeast/yeast.data"),col.names = col_names)
> test1 =aov(datr$nuc ~ datr$class , datr)
> summary(test1)
             Df Sum Sq Mean Sq F value Pr(>F)
datr$class    9  1.993 0.22141   22.01 <2e-16 ***
Residuals  1474 14.825 0.01006
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Solution

Ho = There is no relationship between the discriminant analysis scores of nuclear localization signals of both nuclear and non-nuclear proteins

H1 = There is a relationship between the discriminant analysis scores of nuclear localization signals of both nuclear and non-nuclear proteins

Findings

Based on the p-Value, We fail to reject the null hypothesis and conclude that there is no relationship between the discriminant analysis scores of nuclear localization signals of both nuclear and non-nuclear proteins

b.

ERL is significantly different from the other classes