Part A

H_0: The mean discriminant analysis scores of nuclear localization signals of both nuclear and non-nuclear proteins are equal for all different class type.

H_1: At least one of the class type has different mean discriminant analysis scores of nuclear localization signals of both nuclear and non-nuclear proteins.

Testing at 5% significant level

Conclusion

Since p-value is much smaller than the 5%, hence we reject H_0 and conclude that at least of the class type has its mean discriminant analysis scores of nuclear localization signals of both nuclear and non-nuclear proteins not equal to others.

Used the R code below.

```
> #PART A
> yeast_data = read.csv(file = 'yeast.csv',sep = '',col.names = c('name','
gvh',
+                       'mcg','alm','mit','erl','pox','vac','nuc','class'))
> str(yeast_data)
'data.frame':  1483 obs. of  10 variables:
 $ name : Factor w/ 1461 levels "6P2K_YEAST","6PGD_YEAST",..: 33 34 3 5 4
6 100 7 8 9 ...
 $ gvh  : num  0.43 0.64 0.58 0.42 0.51 0.5 0.48 0.55 0.4 0.43 ...
 $ mcg  : num  0.67 0.62 0.44 0.44 0.4 0.54 0.45 0.5 0.39 0.39 ...
 $ alm  : num  0.48 0.49 0.57 0.48 0.56 0.48 0.59 0.66 0.6 0.54 ...
 $ mit  : num  0.27 0.15 0.13 0.54 0.17 0.65 0.2 0.36 0.15 0.21 ...
 $ erl  : num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
 $ pox  : num  0 0 0 0 0.5 0 0 0 0 0 ...
 $ vac  : num  0.53 0.53 0.54 0.48 0.49 0.53 0.58 0.49 0.58 0.53 ...
 $ nuc  : num  0.22 0.22 0.22 0.22 0.22 0.22 0.34 0.22 0.3 0.27 ...
 $ class: Factor w/ 10 levels "CYT","ERL","EXC",..: 7 7 8 7 1 7 8 7 1 8 ..
.
>
> #Ho: The mean discriminant analysis scores of nuclear localization signa
ls
> #of both nuclear and non-nuclear proteins are equal for all different cl
ass type.
>
> #H1: At least one of the class type has different mean discriminant anal
ysis
> #scores of nuclear localization signals of both nuclear and non-nuclear
proteins.
>
> levels(yeast_data$class)
 [1] "CYT" "ERL" "EXC" "ME1" "ME2" "ME3" "MIT" "NUC" "POX" "VAC"
>
> mean(x = yeast_data$nuc)
[1] 0.2762374
>
> res.aov <- aov(nuc ~ class, data = yeast_data)
>
> # Summary of the analysis
> summary(res.aov)
             Df Sum Sq Mean Sq F value Pr(>F)
class         9   1.99 0.22110   21.97 <2e-16 ***
Residuals  1473  14.82 0.01006
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>  #conclusion
> # P-value is much smaller than the signifance level ,  hence we reject H
o and
> #conclude that at least of the class type has it's mean discriminant ana
lysis
> #scores of nuclear localization signals of both nuclear and non-nuclear
proteins not
> #equal to others
```

Part B

To identify the class type significant different from others let’s use a box plot to plot distribution of the class types.

Looking at the distribution of different class types, the nuc class is significant different from the rest.

```
#PART B
> library(ggplot2)
>
> #Plotting boxplots of different classtypes distributions
> ggplot(yeast_data, aes(x = yeast_data$class, y = yeast_data$nuc)) +
+   geom_boxplot(fill = "grey80", colour = "blue") +
+   scale_x_discrete() + xlab("class types") +
+   ylab("nuc distribution")
> #Looking at the distribution of different class types, the nuc class is
significant
> #different from the rest.
```