Advanced Statistics Practice

# Problem 1:

Read and Solve problem 12.9 from the scanned pdf file titled “Discriminant Analysis Problems”. The data set is labeled as B School.

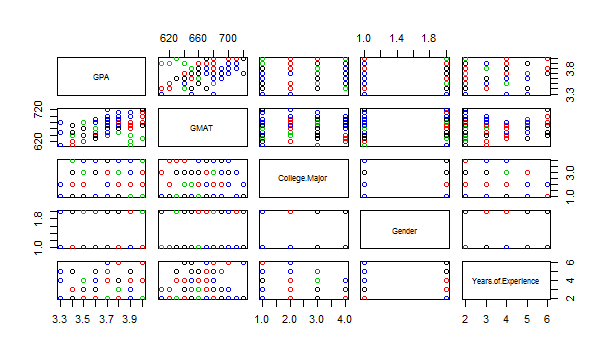
Apply appropriate technique and conduct necessary pre and post tests and report your findings

## Question 1.A

Are the four schools significantly different with respect to the applicants they attract? If so, in what ways they are different?

## Answer 1. A

Below graph tells us that the four schools are not significantly different from each other when checked across all the parameters.



Also, the group means suggests that there is not much difference across the groups as seen below:

Group means:

GPA GMAT CollegeMajor Gender YearsofExperience

1 3.792308 663.0769 2.230769 1.230769 2.461538

2 3.771429 665.7143 2.285714 1.214286 3.285714

3 3.700000 649.0000 2.300000 1.400000 2.500000

4 3.753846 689.2308 2.307692 1.538462 2.538462

However, when a t-Test is ran for the combination of GMAT and B-School type, it shows there is a statistically significant difference in the means across the groups.

## Question 1.B

Another student, from outside the sample, has the following characteristics:  
 Grades:  
 GMAT  
 College Major  
 Gender  
 Years of Experience

What is your best guess, based on your analysis in part a, as to which school this graduate attended?

## Answer 1.B

Based on the model created, it predicts that the student would belong to Business School B from East Coast.

## Question 1.C

How well are you able to discriminate among the graduates of different schools using the information available in the study? Explain briefly.

## Answer 1.C

R-Program used is here:



# Problem 2

Read and Solve problem 12.14 from the scanned pdf file titled “Discriminant Analysis Problems”. The data set is labeled as family car.

What is cross validation? Conduct cross validation and report your classification results.

Apply appropriate technique and conduct necessary pre and post tests and report your findings

## Question 2.A

Does vehicle ownership differ with different household characteristics?

## Answer 2.A

The vehicle ownership differs across various household characteristic.

* The families having car as a vehicle have a higher income when compared to families owning a van. The mean income is highest for the families owning both a car and a van.
* The average family size of van owning families is around 5 while for the ones owning car is around 3.
* The age of household head in car owning families is close to 43 when compared to van owning families which is younger and close to 35.

## Question 2.B

How well does your model perform in using family characteristics to correctly categorize vehicle ownership?

## Answer 2.B

Following table shows the confusion matrix:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **0** | **1** | **2** | **True Positives** |
| **0** | 4 | 0 | 2 | 66.67% |
| **1** | 5 | 2 | 1 | 25.00% |
| **2** | 1 | 0 | 3 | 75.00% |
| **False Positives** | 60.00% | 0.00% | 50.00% |  |

The overall accuracy of the model is 50%.

R program used is here:



# Problem 3

Read and Solve problem 12.10 from the scanned pdf file titled “Discriminant Analysis Problems”. The data set is labeled as Magazines.

Apply appropriate technique and conduct necessary pre and post tests and report your findings

## Question 3.A

Since the result of BoxM test was significant, hence we went for QDA. For QDA we had to remove the non-significant variables and we were left with only "Income", "No\_Male\_Head" and "Education\_Head".

Following are the results of the QDA:

Prior probabilities of groups:

1 2 3 4

0.2428571 0.2428571 0.3000000 0.2142857

Group means:

Income No\_Male\_Head Education\_Head

1 7.000000 0.05882353 5.176471

2 6.176471 0.23529412 5.176471

3 5.952381 0.28571429 4.714286

4 8.333333 0.40000000 6.133333

From the above it seems that the families which have subscribed to Newsweek have a higher Income and the household head on an average has completed a college.

Families with least income averages are opting for TV Guide.

R Program used is here:

