Discriminant Analysis for Correlated Data

Ahmed Mohamed Mohamed El-Sayed

Al-Obour High Institute For Management & Informatics
Department of Basic Science
Kilo 21 Cairo-Belbies Road, P.O. Box 27 Obour City, Egypt
atabl18@yahoo.com - atabl@oi.edu.eg

Abstract

The correlated data are of great importance in the practical life, for example we may want to track the case of a patient after taking a treatment for consecutive periods of time. Also, we may want to track a disease with the members of a certain family. Finally, we want to know the developement of a certain disease with a patient for different periods of time, … , etc. the binary data, (for example a person smokes = 1, a person does not smoke = 0), (female = 1, male = 0), (the effect of treatment is active =1, placebo = 0), is a type of the correlated data. In this paper, we are applying the classification and discriminant methods on the correlated data (some of them are binary data and the other are not). Usually we are using a one dependent variable and then we classify it into two or more classes. Contrary to what is traditionally followed, in this paper we will deal with two dependent correlated categorical variables each one is divided into two classes. The methods of discriminant analysis are investigated in this paper when we are dealing with the independent and dependent variables both of them is correlated variables. Three methods of discriminant analysis are presented for these correlated data. Finally, different packages of R program are used to classify and discriminant the practical data. The data are named respiratory disorder data, these data are containing binary and continous corelated variables. We have applied different methods of classification and discriminant analysis on these data.

Keywords: Linear Discriminant Analysis LDA; Quadratic Discriminant Analysis QDA; Canonical Discriminant Analysis CDA; Logistic Regression LR; Missclassification; Prior Probabilities; Apparent Errors.

1 Introduction

This section presents some definitions and review the classification and discriminant analysis methods. The discriminant analysis is used to predict the categorical dependent variables. It is near to analysis of variance and regression. The discriminant analysis was developed by Fisher [7]. The classification and dicriminant analysis is often presented for one categorical dependent variable and more than one independent continuous variable [6]. The LDA is assumed that the measurements from each class are normally distributed whereas the QDA has no assumption.

The CCA package in R program provides a set of functions that extend the cancor() function. It is used to identify and measure the associations between two sets of quantitative variables. Also, it includes a regularized extension of the canonical correlation [8]. The function (candiscList) generalized CDA for all terms in a multivariate linear model; computing canonical scores and vectors [9]. In the past the CDA is restricted to a one way MANOVA [10]. The package (candisc) is generalized CDA for one term in a multivariate linear model computing canonical scores and vectors [1]. It represents a transformation of the original variables into a canonical space of maximal differences [3]. These relations could be useful in other multivariate data [11]. The LR or logit regression [5] is a regression model where the dependent variable is categorical. The LR was developed by Cox [4]. Re-substitution method evaluates the misclassifications and the cross-validation method estimates the error rate by dividing the data into a training set and the remaining data is called the test data set [2].
The lda() function in R is used for cross-validation method. The apparent error rate method is easier method that uses the training sample set. Then we can calculate the proportions of misclassification for sample units into incorrect population. The classification table is formed by comparing the predicted group for each observation to its actual group. Finally, we will use two methods for the stepwise selection: the backward method; the forward method. Stepwise can be done using the package (klaR) with the function stepclass().

This paper presents a new vision for classifying and discriminat analysis for the respiratory disorder data as a practical correlated data. Since in the past we was using a one dependent categorical variable and then we classify it into two or more groups.
In this paper, contrary to what is traditionally followed, we will deal with two dependent correlated categorical variables (Treatment and Sex) each one is divided into two groups (Active, Placepo) and (Female, Male) respectively. Section 2 presents, numerically, the classification and discriminate analysis for these practical data using three methods of discriminant analysis (LDA, QDA and CDA) using some packages of R program. Section 3 presents some conclusions.

2 Numerical Example

We will classify and discriminant the Respiratory Disorder Data [12]. Some packages of R program like klaR, candisc, Ecdat, CCA,…., etc. will be used in this case. The data present the effect of two treatments on a respiratory disorder illness. It contains 111 patients and 444 observations and 8 variables:

Outcome	: Respiratory status (good=1, poor=0).
Center	: Center (1) = 1, Center (2) = 2.
Id	: Repetition
Age	: Age at time of entry into the study.
Baseline	: Baseline respiratory status (good = 1, poor = 0).
Treatment	: Placebo (P) = 0, Active (A) = 1.
Sex	: Male = 0, Female = 1.
Visit	: Visits [1,2,3 and 4].

We will use the four correlated independent variables (X_1, X_2, X_3 and X_4) to represent the variables (Age, Baseline, Visit and Outcome) respectively.
To study the discriminant analysis of these data, we must have two dependent correlated categorical variables (Sex and Treatment). Each variable has two groups (Sex ={F, M} and Treatment={P, A}). The correlated dependent categorical variables are represented by the variables (Y_1, Y_2) represented the two dependent variables (Sex and Treatment) respectively. The prior probability for each category equals 0.25. Then we can put all variables (two dependents and four independents) in one frame and preceding the classification discriminant analysis using three methods LDA, QDA and CDA.
We are selected two samples: the first sample called the training sample is containing the categories [A, F, M, P] and the second sample (complementary sample) is containing the same categories [A, F, M, P]. Hence each category in the two samples (sets) has [111] observations [A=111, F=111, M=111, P=111] in total [444] observations. The next subsections explain numerically three types of discriminate analysis. Before starting the discriminant analysis, we will present the scatter plot of the training sample (set) in Figure 1:

Figure 1: Training sample

Also, Figure 2 explains the scatter plot of the test sample

Figure 2: Test sample

Since we have shown, from Figures 1-2, the paired correlation between all independent variables (Age,Baseline,Visit,Oucome). Also, the correlation between each independent variable and each category (A, F, M, P) in the train and test sets. Also, we have shown the classification processes for each independent variable according to each category of the two dependent variables (Treatment, Sex), this is clear in the curve distribution for each independent variable. Finally, each group has the same number observations in the train and test sets as shown from the graphical columns for the variables Y's.

The next subsections will present the classification and discriminat processes using the LDA, the QDA and the CDA method.

2.1 The LDA

The partimat() function in the [klaR] package can display the results of linear-quadratic classification using any two variables. In Figure 3 we should expect which independent variable is not contributed to the classification process. The independent variable [Visit] X_3 not contributed with Error = 0.649.

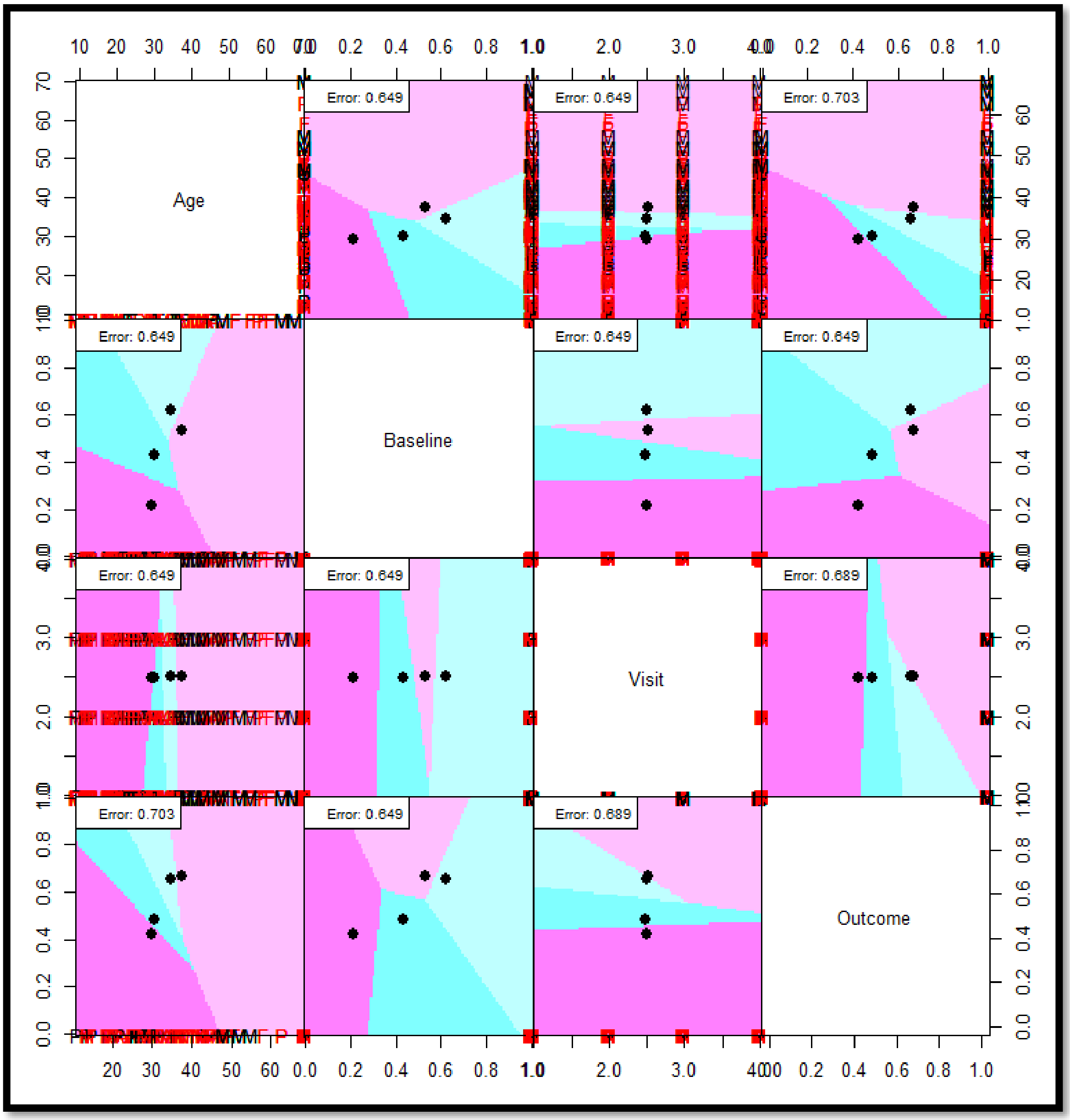

Figure 3: Partimat plot (lda)

Now we will explain the discriminate process for the train and the test samples. We will create a dataset called (math) that has the (respiratory) dataset. We will split the data using the (sample) function. The training dataset will be called (math.train) and the test dataset will be called (math.test). Now we will make our model and it is called (lda.math) and it will include all available variables in the (math.train) dataset.

Next, we will check the results by calling the model (lda.math) that gives us the details of our model. It starts be indicating the prior probabilities 0.25 of someone being (A, F, M or P). We will now use the (predict) function on the train set data to see how well our model classifies the respondents by class (A, F, M or P). We will then compare the prediction of the model with the actual classification. Next, the misclassification table and its probabilities, and the means of each group of the dependent variables, the coefficients of linear dicriminants can be computed for the train.

The missclassification of training sample is explained in Figure 4 as following:

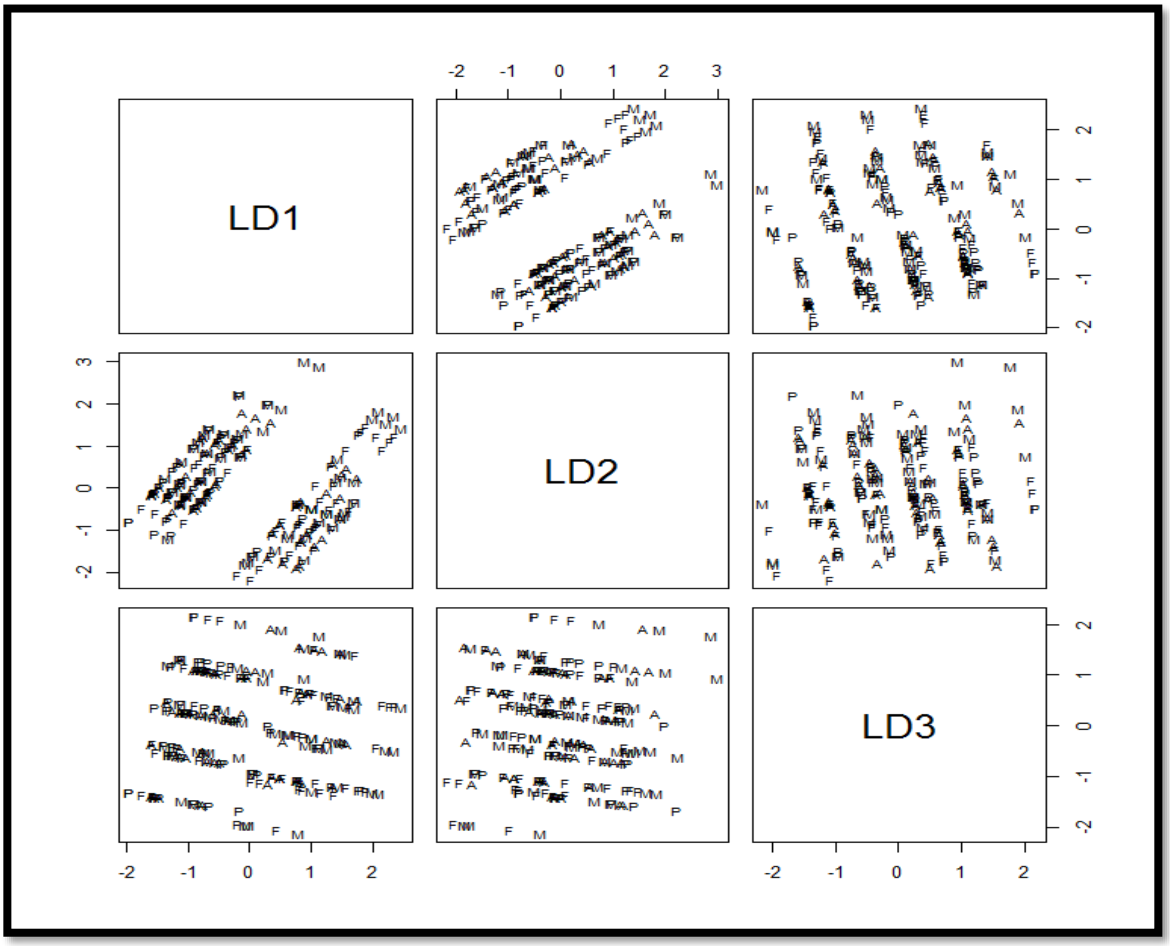

Figure 4: Misclassification of the training sample

Now we have the next results of the training sample:

Prior probabilities:

A	F	M	P
0.2792793	0.2432432	0.2522523	0.2252252

Group means:

	Age	Baseline	Visit	Outcome
A	31.85484	0.3709677	2.564516	0.4838710
F	32.53704	0.6296296	2.500000	0.5555556
M	38.58929	0.5178571	2.464286	0.5892857
P	29.80000	0.1600000	2.340000	0.3600000

Coefficients:

	LD1	LD2	LD3
Age	0.03729653	0.06256106	-0.007236117
Baseline	1.73348119	-1.27369391	-0.610551237
Visit	0.22779943	-0.11507806	0.838159863
Outcome	0.41010383	0.70131482	0.968998562

Correct and Misclassification counts:

	A	F	M	P
A	15	11	10	14
F	18	24	20	7
M	9	10	16	6
P	20	9	10	23

This means that: using the train data for the group [A] we have 15 objects is classified correctly as [A], 18 objects is misclassified as [F], 9 objects is misclassified as [M] and 20 objects is misclassified as [P]. Also, for the group [F] we have 24 objects is classified correctly as [F], 11 objects is misclassified as [A], 10 objects is misclassified as [M] and 9 objects is misclassified as [P]. Also, for the group [M] we have 16 objects is classified correctly as [M], 10 objects is misclassified as [A], 20 objects is misclassified as [F] and 10 objects is misclassified as [P]. Finally, for the group [P] we have 23 objects is classified correctly as [P], 14 objects is misclassified as [A], 7 objects is misclassified as [F] and 6 objects is misclassified as [M].
Dividing each cell of the correct and misclassification counts over the total observations of the train sample gives the proportion of correctness:

Proportion of correctness:

	A	F	M	P
A	0.06756757	0.04954955	0.04504505	0.06306306
F	0.08108108	0.10810811	0.09009009	0.03153153
M	0.04054054	0.04504505	0.07207207	0.02702703
P	0.09009009	0.04054054	0.04504505	0.10360360

Total proportion of correctness = 0.3513514.
Total error rate = 1 - 0.3513514 = 0.6486486.

Also, the missclassification of test sample is explained in Figure 5 as following:

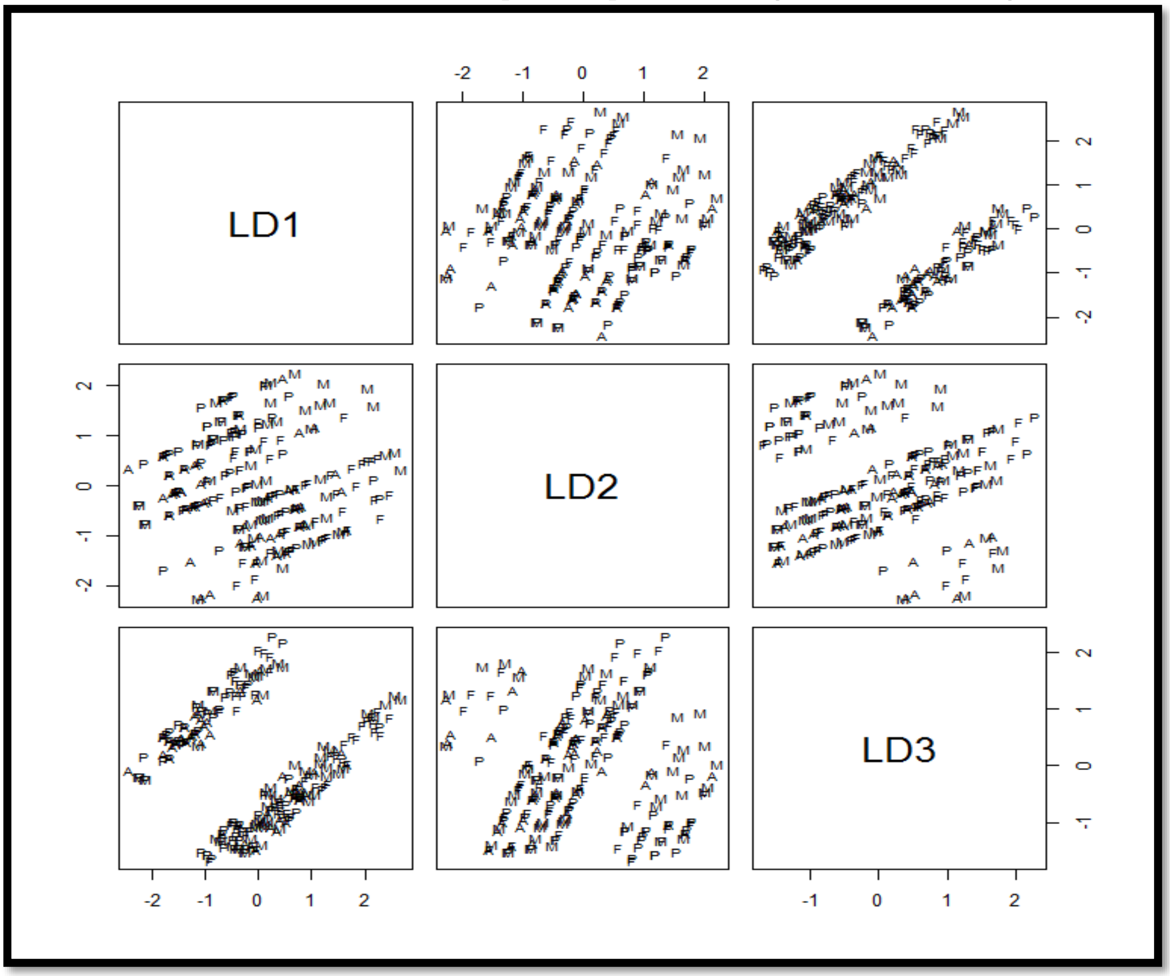

Figure 5: Misclassification of the test sample

Prior probabilities:

A	F	M	P
0.2207207	0.2567568	0.2477477	0.2747748

Group means:

	Age	Baseline	Visit	Outcome
A	29.24490	0.5102041	2.387755	0.4897959
F	36.98246	0.6140351	2.508772	0.7543860
M	36.50909	0.5454545	2.563636	0.7454545
P	30.22951	0.2622951	2.622951	0.4754098

Coefficients:

	Ld-1	Ld-2	Ld-3
Age	0.05491764	0.02107872	0.04816033
Baseline	0.82090784	-1.96757438	0.30366917
Visit	-0.10541749	0.36414559	0.05582148
Outcome	1.30535355	1.20612144	-1.49151593

Correct and Misclassification counts:

	A	F	M	P
A	6	8	9	17
F	22	21	22	13
M	6	18	19	7
P	15	10	5	24

This means that: using the test data for the group [A] we have 6 objects is classified correctly as [A], 22 objects is misclassified as [F], 6 objects is misclassified as [M] and 15 objects is misclassified as [P]. Also, for the group [F] we have 21 objects is classified correctly as [F], 8 objects is misclassified as [A], 18 objects is misclassified as [M] and 10 objects is misclassified as [P]. Also, for the group [M] we have 19 objects is classified correctly as [M], 9 objects is misclassified as [A], 22 objects is misclassified as [F] and 5 objects is misclassified as [P]. Finally, for the group [P] we have 24 objects is classified correctly as [P], 17 objects is misclassified as [A], 13 objects is misclassified as [F] and 7 objects is misclassified as [M].

Proportion of correctness:

	A	F	M	P
A	0.02702703	0.03603604	0.04054054	0.07657658
F	0.09909910	0.09459459	0.09909910	0.05855856
M	0.02702703	0.08108108	0.08558559	0.03153153
P	0.06756757	0.04504505	0.02252252	0.10810811

Total proportion of correctness = 0.3153153.
Total error rate = 1 - 0.3153153 = 0.6846847.

As we can see from the analysis of training and test sets the results are slightly closed. The misclassification error is low within the training set. So, the classification based on the training set is good. The total error rate is slightly higher in the test data (0.6846847), rather training data (0.6486486). The main reason is there is little distinction between the groups (A, F, M and P). This indicates there is a lot of misclassification.

To determine which of the (independent) variables are more affected, we must eliminate the variables that do not improve the correct classification. For this reason we are used two methods.

Figure 6 explains the Backward direction method:

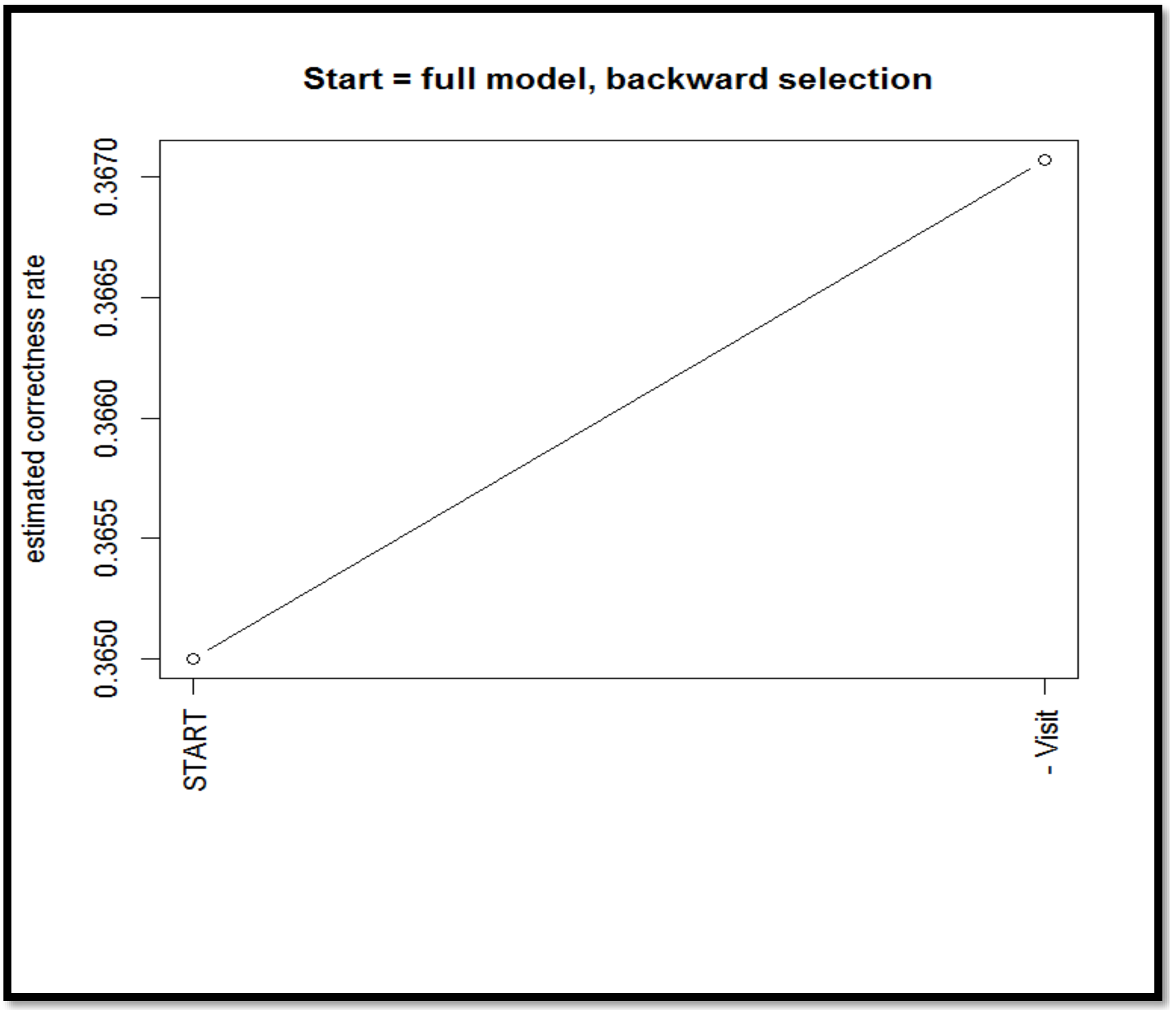

Figure 6 : Backward method

Starting variables (4): Age, Baseline, Visit, Outcome. Correctness rate: [0.365].
Out: Visit. Variables (3): Age, Baseline, Outcome. Correctness rate: [0.36707].
Sec.elapsed = [0.48].
Final model : y ~ Age + Baseline + Outcome

Prior probabilities:

A	F	M	P
0.25	0.25	0.25	0.25

Group means:

	Age	Baseline	Outcome
A	30.70270	0.4324324	0.4864865
F	34.81982	0.6216216	0.6576577
M	37.55856	0.5315315	0.6666667
P	30.03604	0.2162162	0.4234234

Coefficients:

	LD1	LD2	LD3
Age	0.04737358	0.04591772	-0.03768141
Baseline	1.37514908	-1.71761665	-0.56589130
Outcome	0.74583249	1.10632479	1.78912304

As we see from the stepwise [backward] method, we eliminate the variable (Visit) X_3 from the full model which contains four variables [X_1, X_2, X_3 and X_4].
This is because the correctness rate is increased from [0.365] to [0.36707] as we shown in Figure 6. This also explains the same results from a partimat plot in Figure 3.

Figure 7 explains the Forward direction method:

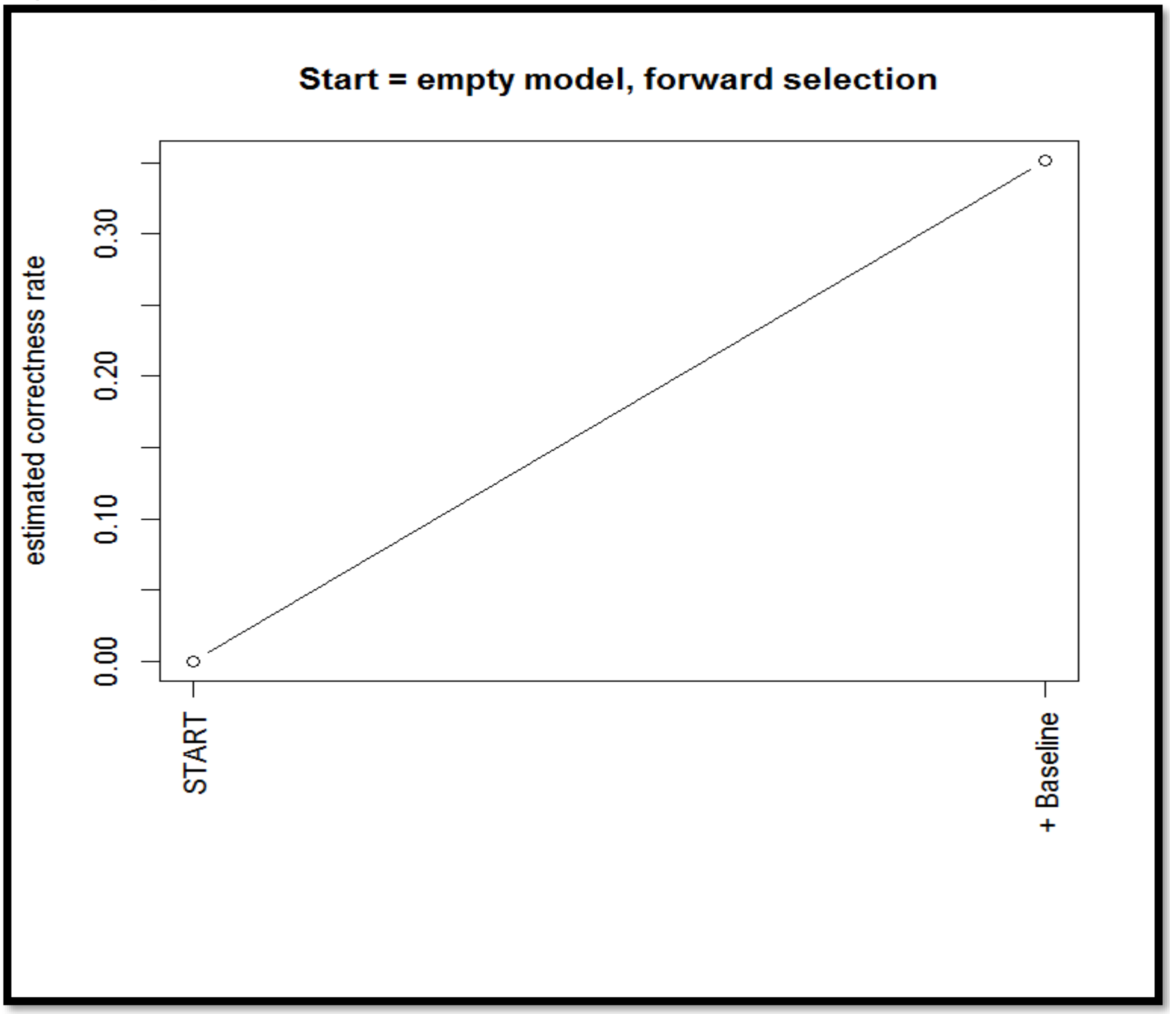

Figure 7 : Forward method

In: Baseline. Variables (1): Baseline. Correctness rate: [0.35141].
Sec.elapsed = [0.36].

Final model : y ~ Baseline.

Prior probabilities:

A	F	M	P
0.25	0.25	0.25	0.25

Group means:

	Baseline
A	0.4324324
F	0.6216216
M	0.5315315
P	0.2162162

Coefficients:

	LD1
Baseline	2.0997

As we see from the above analysis, we added the variable(Baseline) X_2 into an empty model. This is increase the correctness rate [0] to [0.35141] as we shown in Figure 7.

The next subsection presents the QDA as a second method of discriminant analysis of the respiratory disorder data.

2.2 The QDA

We have reviewed the LDA as well as learned about the use of the QDA. Both of these statistical tools are used for predicting categorical dependent variables. The LDA assumes shared covariance in the dependent variable categories but the QDA allows for each category in the dependent variable to have its own variance. Furthermore the QDA does not assume homogeneity of variance-covariance matrices. However, we will see that the QDA will do better in this case. This allows for quadratic terms in the development of the model. We need to use the (qda) function for the training and test data sets. First we will begin with the training set.

The independent variable (Visit) X_3 did not contribute much to the classification process as shown in Figure 8. Since Error = 0.696. This indicates the results that obtained from the LDA.

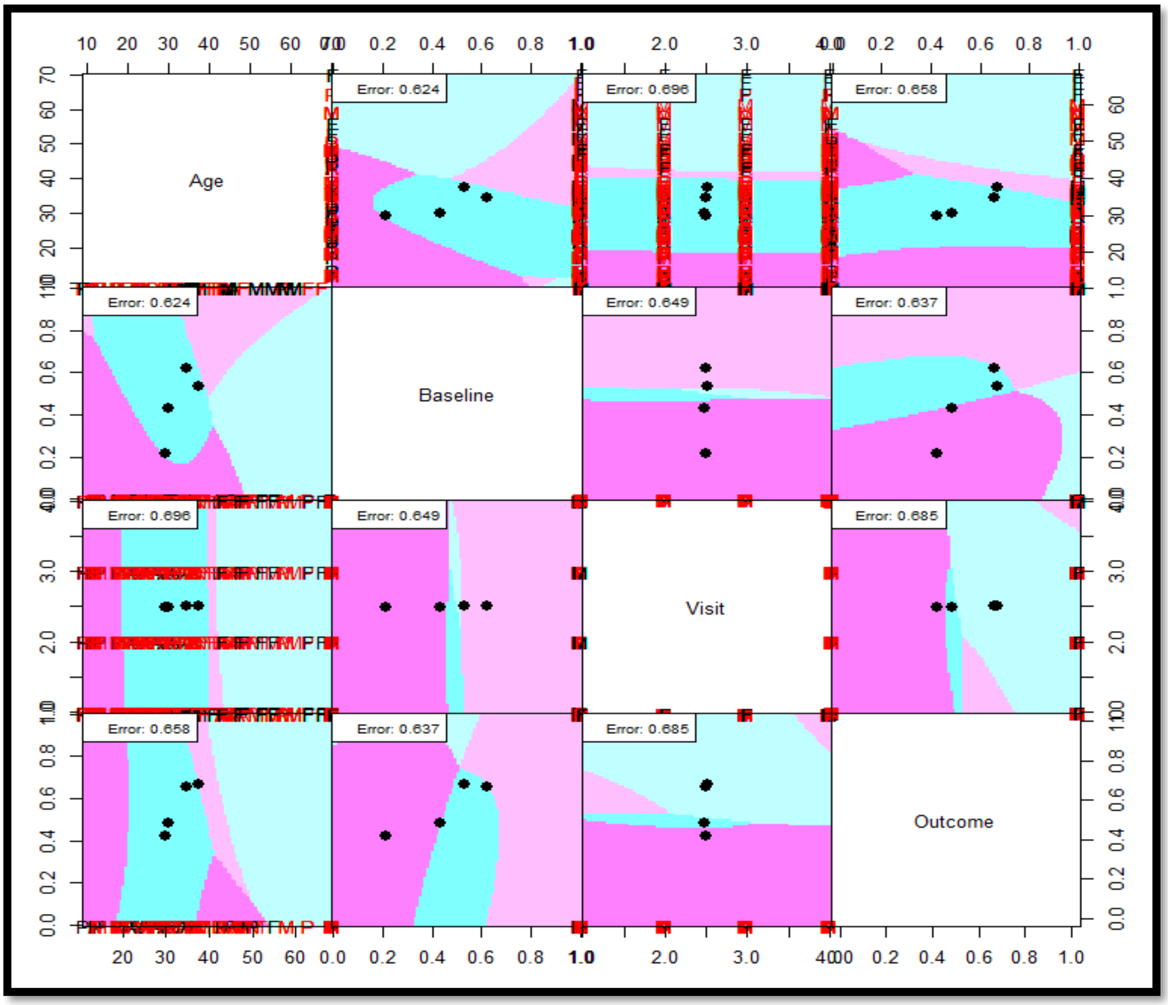

Figure 8: Partimat plot (qda)

As we worked in LDA, we will work in QDA.

The results that obtained from the training sample are:

Prior probabilities:

A	F	M	P
0.2175926	0.2592593	0.2777778	0.2453704

Group means:

	Age	Baseline	Visit	Outcome
A	29.78723	0.4893617	2.446809	0.5319149
F	37.85714	0.4642857	2.553571	0.5892857
M	35.35000	0.6666667	2.433333	0.7166667
P	30.56604	0.2075472	2.433962	0.4150943

Correct and Misclassification counts:

	A	F	M	P
A	13	5	10	6
F	4	17	10	6
M	10	16	27	2
P	20	18	13	39

This means that: using the train data for the group [A] we have 13 objects is classified correctly as [A], 4 objects is misclassified as [F], 10 objects is misclassified as [M] and 20 objects is misclassified as [P]. Also, for the group [F] we have 17 objects is classified correctly as [F], 5 objects is misclassified as [A], 16 objects is misclassified as [M] and 18 objects is misclassified as [P]. Also, for the group [M] we have 27 objects is classified correctly as [M], 10 objects is misclassified as [A], 10 objects is misclassified as [F] and 13 objects is misclassified as [P]. Finally, for the group [P] we have 39 objects is classified correctly as [P], 6 objects is misclassified as [A], 6 objects is misclassified as [F] and 2 objects is misclassified as [M].

Dividing each cell of the correct and misclassification counts over the total observations of the train sample gives the proportion of correctness as shown:

Proportion of correctness:

	A	F	M	P
A	0.060185185	0.023148148	0.046296296	0.027777778
F	0.018518519	0.078703704	0.046296296	0.027777778
M	0.046296296	0.074074074	0.125000000	0.009259259
P	0.092592593	0.083333333	0.060185185	0.180555556

Total proportion of correctness = 0.4444444.
Total error rate = 1 - 0.4444444 = 0.5555556.
As we know we did not have linear discriminants because we are using the QDA.

The results that obtained from the test sample are:

Prior probabilities:

A	F	M	P
0.2807018	0.2412281	0.2236842	0.2543860

Group means:

	Age	Baseline	Visit	Outcome
A	31.37500	0.3906250	2.515625	0.4531250
F	37.25455	0.6000000	2.472727	0.7454545
M	34.19608	0.5686275	2.588235	0.5882353
P	29.55172	0.2241379	2.551724	0.4310345

Correct and Misclassification counts:

	A	F	M	P
A	6	9	8	6
F	6	14	7	5
M	17	24	15	6
P	34	8	21	41

This means that: using the test data for the group [A] we have 6 objects is classified correctly as [A], 6 objects is misclassified as [F], 17 objects is misclassified as [M] and 34 objects is misclassified as [P]. Also, for the group [F] we have 14 objects is classified correctly as [F], 9 objects is misclassified as [A], 24 objects is misclassified as [M] and 8 objects is misclassified as [P]. Also, for the group [M] we have 15 objects is classified correctly as [M], 8 objects is misclassified as [A], 7 objects is misclassified as [F] and 21 objects is misclassified as [P].

Finally, for the group [P] we have 41 objects is classified correctly as [P], 6 objects is misclassified as [A], 5 objects is misclassified as [F] and 6 objects is misclassified as [M].

Proportion of correctness:

	A	F	M	P
A	0.02631579	0.03947368	0.03508772	0.02631579
F	0.02631579	0.06140351	0.03070175	0.02192982
M	0.07894737	0.10526316	0.06578947	0.02631579
P	0.14912281	0.03508772	0.09210526	0.17982456

Total proportion of correctness = 0.3333333.
Total error rate = 1 - 0.3333333 = 0.6666667.

Also, the stepwise direction will be done for the classification process using the QDA. The backward direction is explained in Figure 9 as follows:

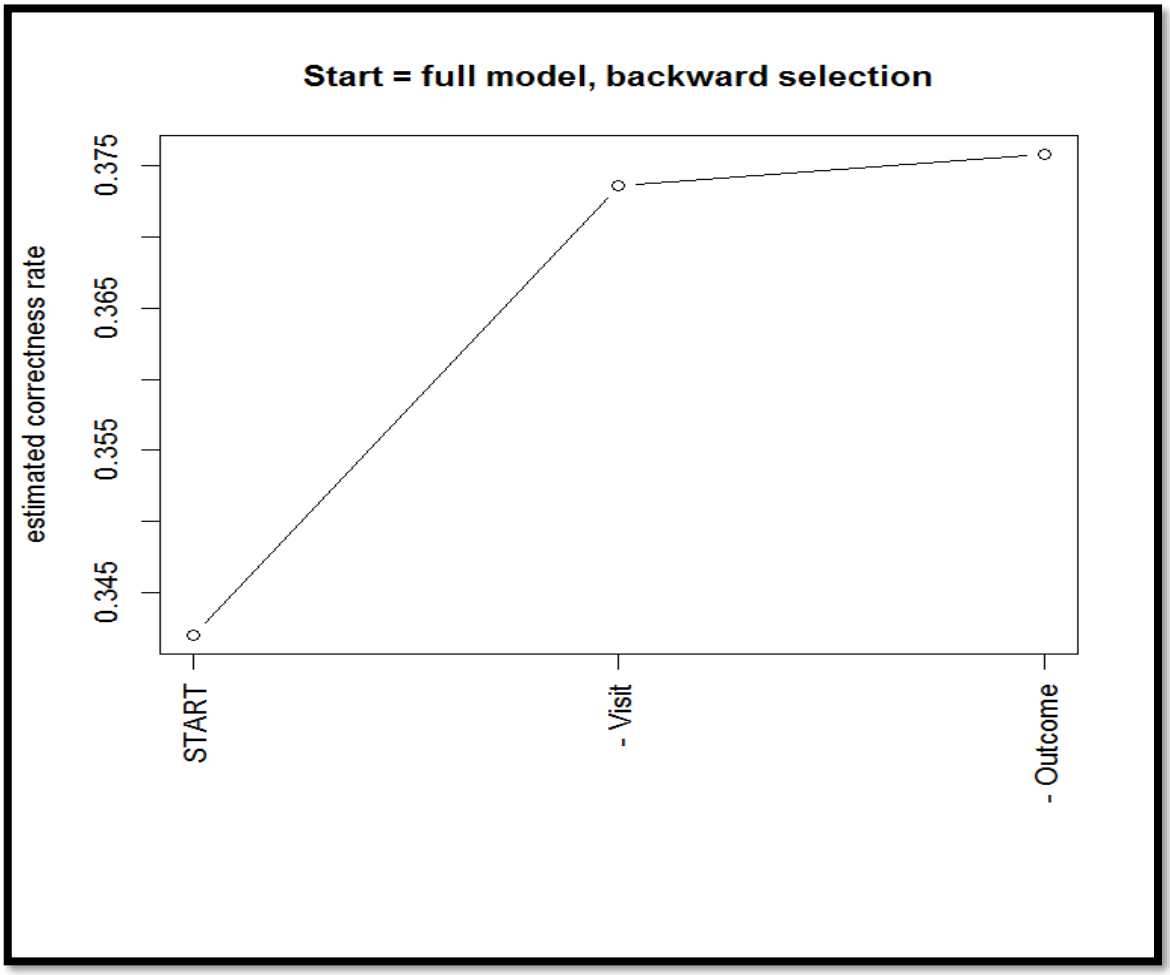

Figure 9 : Backward method

Starting variables (4): Age, Baseline, Visit, Outcome. Correctness rate: [0.34202].
Out: Visit. Variables (3): Age, Baseline, Outcome. Correctness rate: [0.37359].
Out: Outcome. Correctness rate: [0.37581]. Variables (2): Age, Baseline.
Sec.elapsed = [0.81].
Final model : y ~ Age + Baseline.

Prior probabilities:

A	F	M	P
0.25	0.25	0.25	0.25

Group means:

	Age	Baseline
A	30.70270	0.4324324
F	37.55856	0.5315315
M	34.81982	0.6216216
P	30.03604	0.2162162

Figure 10 explains the forward selection method as follows:

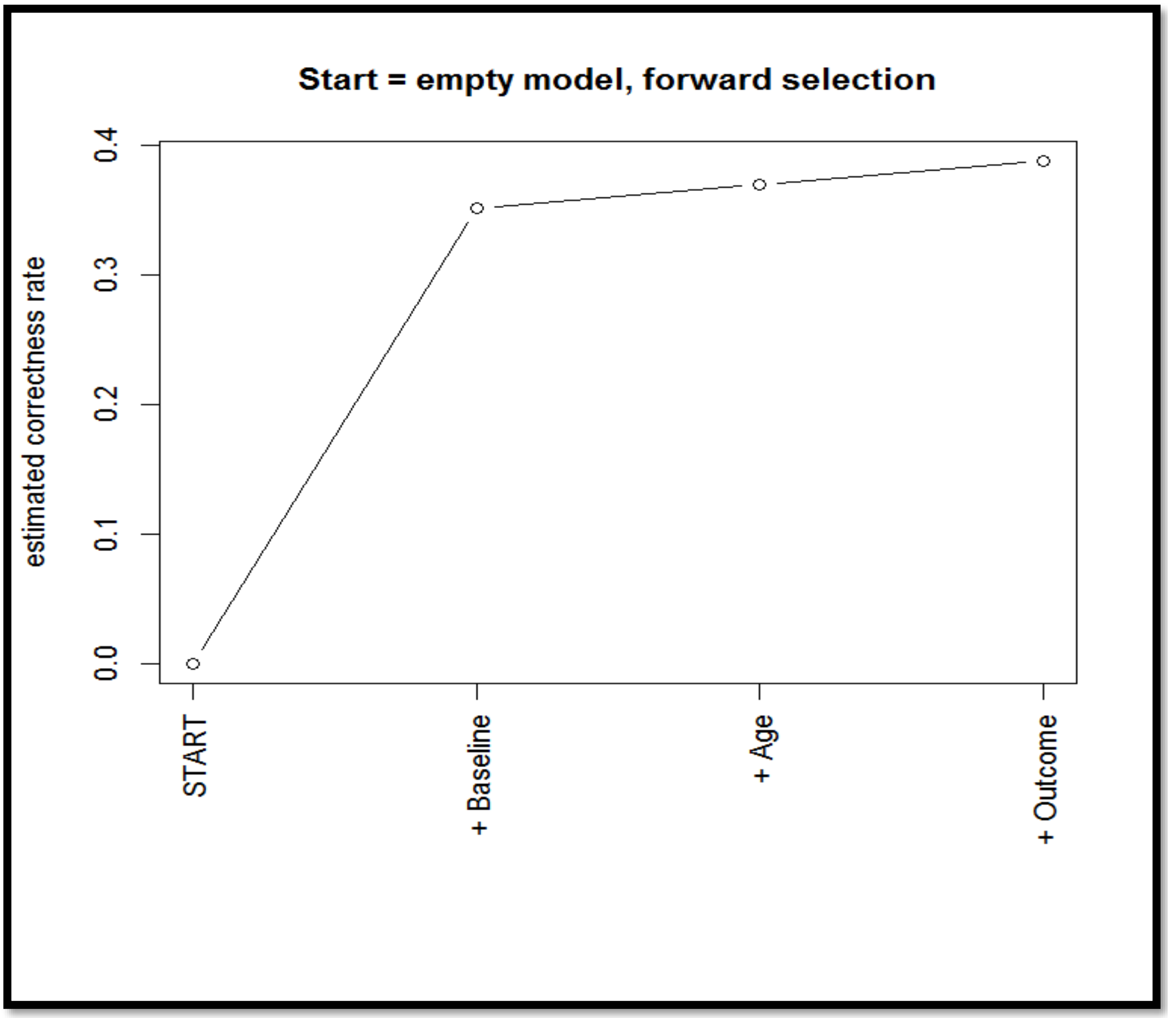

Figure 10 : Forward method

In: Baseline. Variables (1): Baseline. Correctness rate: [0.35126].
In: Age. Variables (2): Baseline, Age. Correctness rate: [0.36934].
In: Outcome. Variables (3): Baseline, Age, Outcome. Correctness rate: [0.38747].
Sec.elapsed = [0.87]
Final model : y ~ Age + Baseline + Outcome.

Prior probabilities:

A	F	M	P
0.25	0.25	0.25	0.25

Group means:

	Age	Baseline	Outcome
A	30.70270	0.4324324	0.4864865
F	37.55856	0.5315315	0.6666667
M	34.81982	0.6216216	0.6576577
P	30.03604	0.2162162	0.4234234

The next subsection presents the CDA as a third method of discriminant analysis of the respiratory disorder data.

2.3 The CDA

In this section, we will investigate the CDA as follow:

The independent variables here are two variables (Treatment and Sex) which have four classes (A, P) and (F, M) for each variable respectively. These variables are correlated. The dependent variables are also correlated variables (Age, Baseline, Visit and Outcome). The CDA has the next formula:

Formula = cbind(Age, Baseline, Visit, Outcome) ~ y.

Figures 11.a - 11.b presents the Canonical scores for these variables:

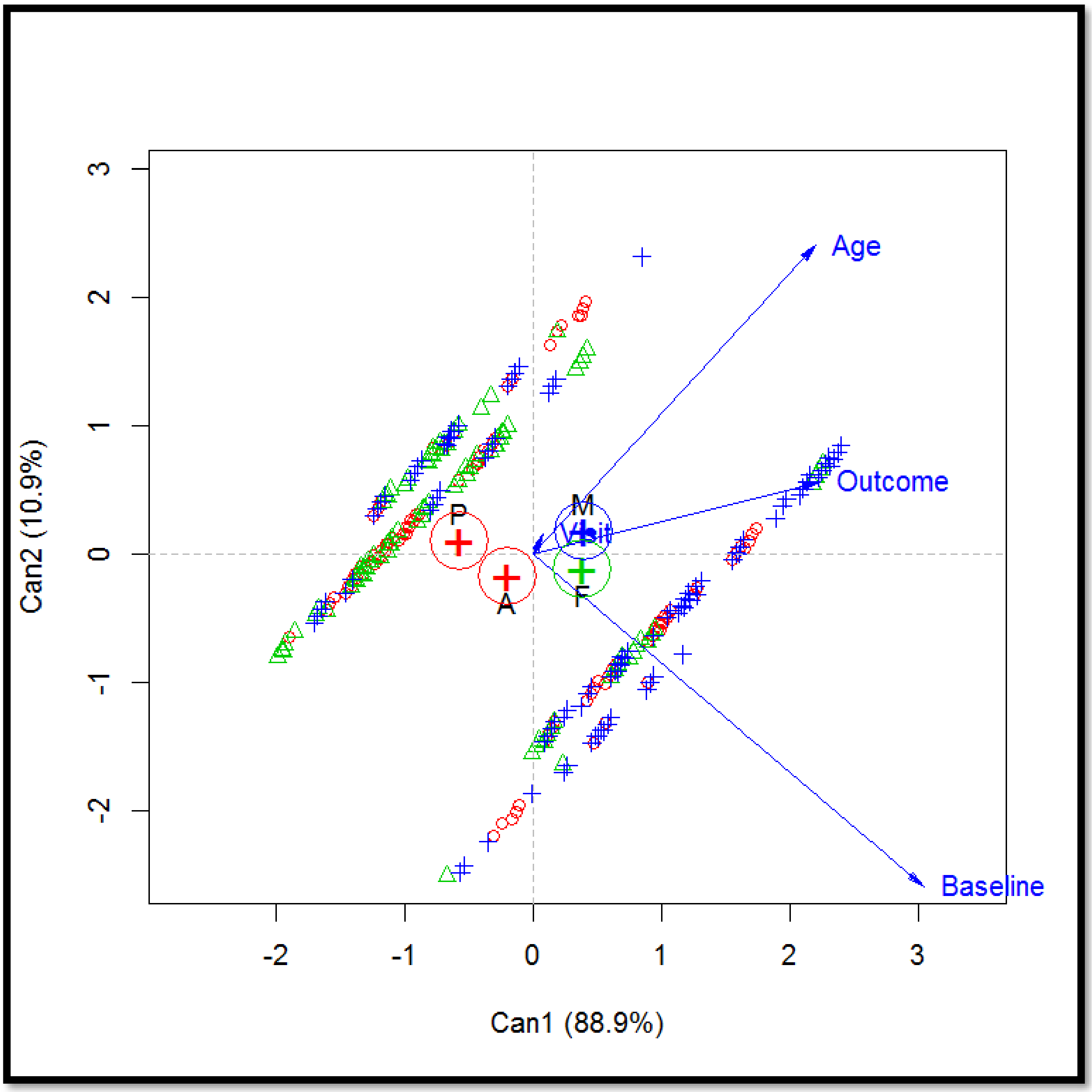

Figure 11.a: Canonical scores

As we shown in Figure 11.a the variable [Visit] arises in the Error region. This insures that it must be exposed from full model. Figure 11.b explains the canonical scores error region for these variables:

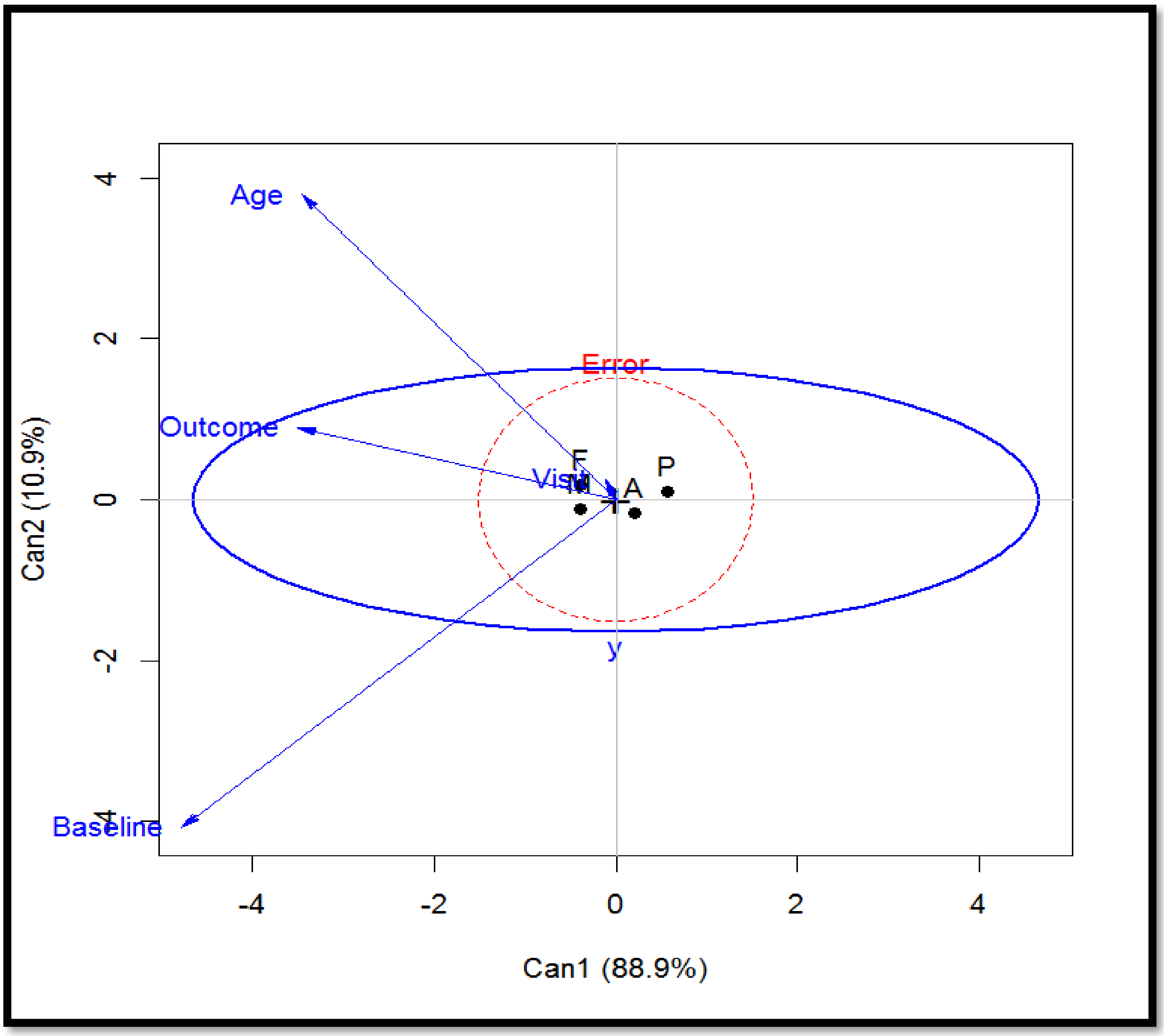

Figure 11.b: Canonical scores error

The linear coefficients of CDA:

Coefficients:

	Age	Baseline	Visit	outcome
A	30.702703	0.432432	2.486486	0.486486
F	6.855856	0.099099	0.027027	0.180180
M	4.117117	0.189189	0.018018	0.171171
P	-0.666667	-0.216216	0.009009	-0.063063

Class means:

	Can1	Can2
A	0.20112	-0.16749
F	-0.39536	0.17692
M	-0.38232	-0.11375
P	0.57655	0.10432

Standardized canonical coefficients:

	Can1	Can2
Age	-0.630081	0.608651
Baseline	-0.653979	-0.818814
Visit	-0.031217	0.058067
Outcome	-0.364690	0.539936

Canonical correlation analysis:

	CanR	CanRSQ	Eigen-value	Percent	Cumulative
A	0.38150	0.1456000	0.1704000	88.88	88.88
F	0.14340	0.0205500	0.0209800	10.95	99.82
M	0.01845	0.0003403	0.0003404	0.1776	100.00
P	5.425e-11	2.943e-21	2.943e-21	1.535e-18	100.00

The test:

CanR	L. Ratio	Approximate F	Numerator df	Denominator df	p-value
0.38153	0.83659	5.0093	16	1332.6	3.452e-10 ***
0.14336	0.97911	1.0295	9	1063.7	0.4140
0.01845	0.99966	0.0373	4	876.0	0.9974
0.00000	1.00000	0.0000	1	439.0	1.0000

Significant codes: *** 0.001

We have a significant canonical correlation in the group [A]. The canonical correlations in the other groups [F, M and P] are zero.

Wilks test:

df =3, test statistic = 0.83659, approximate F = 6.7235 and p-value = 8.597e-12 ***.

We have a significant association between all classes.

3 Conclusions

In this paper, we dealt with the respiratory disorder data. These data have two dependent correlated categorical variables (Treatment and Sex). For the classification and discriminant analysis each one is divided into two groups. This means that we are dealt with four calsses (Active, Placepo, Femal and Male). Also, four independent variables (Age, Baseline, Visit and Outcome) all are correlated variables. Three methods of the discriminant analysis are explained. These methods are LDA, the QDA, and the CDA. Different packages of R program are used to classify and discriminant the practical data. Some tables and graphs are presented for explaination and comparison. Finally, some results are conducted from the previous discussion:

1. For the tarining set: the LDA method has total proportion of corectness (0.3513514) it is less than the QDA method (0.4444444). The total error in the LDA equals (0.648648) it is more than the QDA (0.5555556). This means that the QDA method is better than the LDA method for the classification process.

2. For the test set: the LDA has total proportion of corectness (0.3153153) it is less than the QDA (0.3333333). The total error in the LDA equals (0.6846847) it is more than the QDA (0.6666667). This also means that the QDA method is better than the LDA method for the classification process.

3. For the stepwise (backward) selection: in the LDA, the variable (Vist) is eleminated from the full model, and the correctness rate is raised to (0.36707). But in the QDA, the variables (Visit and Outcome) are eleminated from the full model sequentially, and the corectness rates is raised to (0.37581).

4. For the stepwise (forward) selection: in the LDA, the variable (Baseline) is added to the null model, and the correctness rate is raised to (0.35141). But in the QDA the variables (Baseline, Age and Outcome) are added to the null model sequentially, and the corectness rates is raised to (0.38747).

5. The error rate in all partimat plots indicates that the variable (Visit) was not contributed to other independent variables, with error (0.649) in the LDA method, and with error (0.696) in the QDA method. Also, the CDA method insured that the variable (Visit) must be eleminated from the full model as explained in Figure 11.b. Since the canonical score error region arrounds this variable.

6. The QDA and CDA methods have not linear predictors but the LDA method has.

References

[1] Bartlett, M. S. (1938). Further aspects of the theory of multiple regressions. Proc. Camb. Phil. Soc. 34, pp. 33-34.

[2] Bedrick, E. J., Schrader, R. M. and Erhardt, E. B. (2003). Lecture notes for advanced data analysis 2 (ADA2) Stat 428/528. University of New Mexico.

[3] Cooley, W.W. and Lohnes, P.R. (1971). Multivariate data analysis, Wiley, New York:.

[4] Cox, D.R. (1958). The regression analysis of binary sequences (with discussion). J. Roy Stat Soc. B. 20, pp. 215–242.

[5] David, A. F. (2009). Statistical models: Theory and practice. Cambridge University Press, pp. 128.

[6] Efron, B. and Tibshirani, R. J. (1993). An introduction to the bootstrap. Chapman Hall, New York, USA.

[7] Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. Annals of Eugenics 7, 2, pp.179–188.

[8] Friendly, M. (2007). HE plots for multivariate general linear models. Journal of Computational and Graphical Statistics, 16(2), pp. 421–444.

[9] Friendly, M. and Kwan, E. (2003). Effect ordering for data displays. Computational statistics and data analysis, 43, pp. 509-539.

[10] Friendly, M. and Sigal, M. (2014). Recent advances in visualizing multivariate linear models. Revista Colombiana de Estadistica, 37(2), pp. 261-283.

[11] Gittins, R. (1985). Canonical analysis: A review with applications in ecology, Berlin: Springer.

[12] Stokes, Davis and Koch (1995). SAS and R programs.