International Journal of Transformation in Applied Mathematics & Statistics
Vol. 3, Issue 1 – 2020
ISSN: 2581-7620

Discriminant Analysis of Unemployment and Literacy Rate by State in Nigeria

George, Daniel Sokari[1], Enegesele, Dennis[2], Biu, O. Emmanuel[3], Dagogo, Joseph[3]

[1]*Ignatius Ajuru University of Education, Port Harcourt, Rivers State.*
[2]*Department of Mathematics and Computing Sciences, KolaDaisi University, Ibadan, Nigeria.*
[3]*University of Port Harcourt, Port Harcourt, Rivers State.*

Abstract

This paper used Fisher's linear discriminant functions to classified and determined if unemployment and literacy rate in Nigeria by State differs. The multivariate data covariates investigated are unemployment and literacy rate in Nigeria states. The data were classified using low and high rates. Fisher's linear discriminant functions, Standardized Canonical Discriminant (SCD) Function, and Group Centroids (GC) Function were obtained. The result of Wilks' Lambda statistics show that the equality of group means between low and high unemployment rate is significant at 5% but the literacy rate is not significant. This result indicates that the unemployment rate in Nigeria by state differs significantly while the literacy rates are similar by state. The classification results show that 18 states were correctly classified with a high level of the unemployment rate, while 18 states are accurate with a low unemployment rate. However, only one state was misclassified, using discriminant Score functions p-values.

Keywords: Fisher's linear discriminant functions, Standardized Canonical Discriminant (SCD) Function, Group Centroids (GC) Function, Wilks' Lambda statistics, classification

Introduction

Linear discriminant function analysis (i.e., discriminant analysis) performs a multivariate test of differences between groups and it is used to determine the minimum number of dimensions needed to describe these differences. A distinction is sometimes made between descriptive discriminant analysis and predictive discriminant analysis. In different areas of application, Discriminant Analysis has earned a different name, uses, and roles for itself (Stevens, 2002). It has also been used as an adjunct to MANOVA.

Discriminant analysis is a general technique for analysing data when the criterion or dependent variable is categorical and the predictor or independent variables are intervals in nature. The objectives of discriminant analysis are as follows: 1) development of discriminant functions, or linear combinations of the predictor or independent variables, that best discriminate between the categories of the criterion or dependent variable (Groups: low and high unemployment's); 2) examination of whether significant differences exist among the groups, in terms of the predictor variables (Alayande and Bashiru, 2015); 3) determination of which predictor variables contribute to most of the inter-group differences; 4) classification of cases to one of the groups based on the values of the predictor variables; and 5) evaluation of the accuracy of classification (Huberty and Olejnik, 2006).

The rationale for this paper is the problem classifying the unemployment and literacy rate in the states in Nigeria. Thus, the paper aims to correctly classify the unemployment rate as high and low as well as the literacy rate. The specific objectives will be to identify if unemployment and literacy rate is different in the Nigerian state.

This study is divided into five sections. Section 1 contains the introduction of discriminant analysis, rationale, and aim of the study, Section II contains the literature on Discriminant analysis, Section III described the materials and methods, Section IV described results and discussion, and Section V concludes the research work.

Literature

Discriminant analysis techniques are described by the number of categories possessed by the criterion variable. When the criterion variable has two categories, the technique is known as two-group discriminant analysis. When three or more categories are involved, the technique is referred to as multiple discriminant analysis. The main distinction is that in the two-group case it is possible to derive only one discriminant function, but in multiple discriminant analysis, more than one function may be computed.

Discriminant Analysis is a classic method of classification that has stood the test of time. Discriminant analysis often produces models whose accuracy approaches (and occasionally exceeds) more complex modern methods. Discriminant analysis can be used only for classification (i.e., with a categorical target variable), not for regression (Ramayah, *at al.,* 2010). The target variable may have two or more categorical data. The objective of discriminant analysis is to classify objects, by a set of independent variables, into one of two or more mutually exclusive and exhaustive categories. For example, in this case, based on the unemployment rate and literacy rates, the researcher wants to classify the states into high and low unemployment rates. For notation, let X_{ij} be the i^{th} individual's value of the jth independent variable, b_j be the discriminant coefficient for the j^{th} variable, $K_i$is the i^{th} individual's discriminant score, and k_{erit} be the critical value for the discriminant score. Under the linear classification procedure, let each individual's discriminant score K_i be a linear function of the independent variables. That is

$$K_i = ß_0 + ß_1 X_i + ß_2 X_{2i} + ...+ ß_i X_{ni} \quad (1)$$

The classification procedure follows: if Ki > $K_{critical}$, classify Individual i as belonging to Group 1; if K_i <$K_{critical}$, classify Individual i as belonging to Group 2. The classification boundary will then be the locus of the points, where $ß_0 + ß_1 X_i + ß_2 X_{2i} + ...+ ß_i X_{ni}$.

When n (the number of independent variables) is equal to 2, the classification boundary is a straight line. Every individual on one side of the line is classified as Group 1 and on the other side, as Group 2. When n = 3, the classification boundary is a two-dimensional plane in 3-dimensional space; the classification boundary is generally an n-1 dimensional hyperplane in n space.

Discriminant Analysis is of Different types; they are the Multiple Discriminant Analysis, Linear Discriminant Analysis, and the K-NNs Discriminant Analysis. In this paper, the Linear Discriminant Analysis was adopted.

Discriminant Analysis has the following assumptions

1) Multivariate Normality
2) Data values are from a normal distribution. However, normal assumptions are usually not "fatal". The resultant significance tests may still be reliable.
3) Equality of variance-covariance within the group
4) The covariance matrix within each group should be equal. Equality Test of Covariance Matrices can be used to verify it.

Low multicollinearity of the variables. When high multicollinearity among two or more variables is present, the discriminant function coefficients will not reliably predict group membership. The use of pooled within-groups correlation matrix can be used to detect multicollinearity. If there are correlation coefficients larger than 0.8, some variables excluded or Principle is used (Alayande and Bashiru, 2015).

Materials and Methods

When talking about factors that may have impacted on unemployment rate in Nigeria, one would easily think about crime rates, youth restiveness, turn out of a large number of graduate more than the labour market, location, gender, population density, level of urbanisation, the composition of the populations, and many more. In this paper, unemployment rates by state and literacy levels were used to classify high and low unemployment rates. The data used in his paper is from the National Bureau of Statistics (third quarter of 2018). This form of discriminant analysis is the linear discriminant analysis for two groups only. (See appendix for data details).

The sums of squares and cross products matrices for K^{th} group (Note: K is two in this study), S_T, S_W, and S_E, is defined as follows:

$$S_T = \sum_{k=1}^{K} \sum_{j=1}^{N_k} (X_{Ki} - M)(X_{Ki} - M)' \qquad (2)$$

$$S_W = \sum_{k=1}^{K} \sum_{j=1}^{N_k} (X_{Ki} - M_k)(X_{Ki} - M_k)' \qquad (3)$$

$$S_E = S_T - S_W \qquad (4)$$

where $N_k$is observations per group. Let N represent the total number of observations. Each observation consists of the measurements of p variables. The i^{th} observation is represented by X_{ki}. M represents the vector of means of these variables across all groups and M_k the vector of means of observations in the k_{th} group.

The two degrees of freedoms, df_1 and df_2 is given by; Df_1 = K-1 and Df_2 = N-K.

A discriminant function is a weighted average of the values of the independent variables. The weights are selected so that the resulting weighted average separates the observations into the groups. High values of the average come from one group, low values of the average come from another group. Thus, the problem reduces to one of finding the weights which, when applied to the data, best discriminate among groups according to some criterion. The solution reduces to finding the eigenvectors, V. The canonical coefficients are the elements of these eigenvectors.

A goodness-of-fit parameter, Wilks' lambda, is defined as follows:

$$? = \frac{|S_W|}{|S_T|} = \prod_{j=1}^{m} \left(\frac{1}{1+\lambda_j}\right) \qquad (5)$$

where λ_j is the j^{th} eigenvalue corresponding to the eigenvector described above and m is the minimum of K-1 and p

The canonical correlation between the j^{th} discriminant function and the independent variables is related to these eigenvalues as follows:

$$r_{cj} = \sqrt{\frac{\lambda_j}{1+\lambda_j}} \qquad (6)$$

The overall covariance matrix T, is given by

$$T = \left(\frac{1}{N-1}\right) S_T \qquad (7)$$

The within-group covariance matrix, W, is given by

ISSN: 2581-7620

$$W = \left(\frac{1}{N - K}\right) S_T \tag{8}$$

The among (between) group covariance matrix, A, is given by

$$A = \left(\frac{1}{K - 1}\right) S_A \tag{9}$$

The linear discriminant functions are defined as:

$$\mathrm{LDF}_k = W^{-1} M_k \tag{10}$$

The standardised canonical coefficient is given by

$$V_{ij}\sqrt{W_{ij}} \tag{11}$$

where V_{ij} are the elements of V and W_{ij} are the elements of W

The correlation between the independent variable and the canonical variates are given by

$$\mathrm{Corr}_{jk} = \frac{1}{\sqrt{w}} \sum_{i=1}^{p} y_{ik}\, w_{ji} \tag{12}$$

Results and Discussion

The analysis of the case summary in Table 1 shows that all 37 cases were valid giving a total with 59.7% accuracy in the entry. This entry is for the unemployment rate data collected from the national bureau of statistics online. The data was that of the third quarter of 2018. The authors also observed that there were no cases of exclusion and none was missing as well. However, a holdout value of 25 was indicated. The out of range values are from the high and low entries of the dependent variable which was used for grouping and or classification. Thus, 100% accuracy was ensured.

Table 1.Analysis Case Processing Summary

Unweighted	**Frequency (Percent)**
Valid	37 (59.7%)
Excluded	25 (40.3%)
Total	62(100.0)

Table 2.Descriptive statistics of variables (Mean± Std. Deviation)

Group	**Low**	**High**	**Total**
Unemployed	17.6167±3.5568	28.9842±3.8822	23.4541±6.8329
Literacy	68.3167±11.4138	70.0263±9.8500	69.1946±10.5252
State	20.6111±10.0595	17.4737±11.5632	19.0000±10.8244
t-test statistics	-0.78 (0.518) [95% CI: 21.66, 15.03]		

International Journal of Transformation in Applied Mathematics & Statistics
Vol. 3, Issue 1 – 2020
ISSN: 2581-7620

Table 2 above indicates that the weighted and unweighted valid cases are equal. It further showed that the individual group means and standard deviations differ at both high and low group levels. However, when tested with paired t-test statistics, it's suggested otherwise since the confidence interval for the mean difference between the levels includes zero, which suggests no difference between the levels. The p-value of 0.518 further suggests that the classification is consistent with $H_0:\mu_d = 0$, that is, the two levels perform equally. The probability distributions of covariates are shown in Table 3.

Table 3.Probability distribution of variables

GROUP	LOW	HIGH	P(covariates)
Unemployed	0.08	0.13	0.21
Literacy	0.31	0.31	0.62
State	0.09	0.08	0.17
P(Level)	0.48	0.52	1.00

The probability distribution between the levels in Table 3 shows 48% for low level and 52% for high. Literacy is equal between the levels, while the unemployment rate is slightly high with 13% for high and 8% for low respectively. The state covariate shows 9% low and 8% high. Also, there are more literacy individuals with 62% than unemployed with a 21% rate.

Table 4.Tests of Equality of Group Means

Groups	Wilks' Lambda	F (p-values)
Unemployed	0.289	85.956 (0.000***)
Literacy	0.993	0.239(0.628)
State	0.978	0.772(0.386)

Footnote: ***= sig. at 1%

The Wilks' Lambda statistics used for tests of equality of group means by covariate in Table 4 above shows that the test for equality of group means between low and high rate is significant at unemployment but not in literacy and state.

The obtained Fisher's linear discriminant functions in Table 5 for Classification is shown below;

Table 5.Classification Function Coefficients

D(X)	GROUP	
	LOW	HIGH
Constant	-41.768	-63.925
Unemployed	1.549	2.421
Literacy	0.664	0.673
State	0.462	0.526

Thus, Fisher's linear discriminant functions are:

International Journal of Transformation in Applied Mathematics & Statistics
Vol. 3, Issue 1 – 2020
ISSN: 2581-7620

D(Low) = -41.786+ 1.549(Unemployed)+ 0.664(Literacy)+ 0.462(State) (13)

and

D(High) = -63.925+ 2.421 (Unemployed)+ 0.673 (Literacy)+ 0.526 (State) (14)

The class function coefficient in Table 5 above indicates 1.549 and 2.421 for unemployment low and high rates; 0.664 and 0.673 for literacy low and high, and 0.462 and 0.526 for state low and high rate respectively. The classification function coefficients show a negative constant spooling of -41.768 and -63.925 for low and high as well.

Table 6.Box's Test of Equality of Covariance Matrices (The ranks and natural logarithms of determinants printed are those of the group covariance matrices)

Group	Low	High	Pooled within-groups
Log Determinant	11.672	11.847	11.998
Box's M	8.263		
F-Approx. (p-value)	1.248 (0.278)		
Eigenvalue	2.568[a]		
% of Variance	100.0		
Canonical Correlation	0.848		
Wilks' Lambda	0.280 (0.000***)		
Chi-square (p-value)	42.611 (0.000***)		

Footnote: ***= sig. at 1% and a. First 1 canonical discriminant functions were used in the analysis

The Box's test of equality of Covariance Matrices in Table 6 shows a constant behaviour of the log determinant with 11.672, 11.847, and 11.998 for a low, high, and within groups respectively. The computed eigenvalue, which is the ratio of between-group to within-group sum of square of high and low classification is 2.568, it implies that the States with low rate unemployment are slightly lower than the States with high rate unemployment by approximating 3%. The canonical correlation of 0.848, shows a substantial relationship between the high and low classification. The Wilks' Lambda (0.280) and chi-square (42.611) values with p-values (0.000) were significant at 1%.

Table 7.Standardized Canonical Discriminant (SCD) Function Coefficients and Functions at Group Centroids (GC)

F(X): Covariate	F(SCD)	F(GC)
Unemployed	1.042	0.978
Literacy	0.031	-0.093
State	0.222	0.052

The functions are:

F(SCD) = 1.042(Unemployed) + 0.031(Literacy) + 0.222(State)

F(GC) = 0.978 (Unemployed) - 0.093 (Literacy) + 0.052 (State)

The standardised canonical discriminant and group centroids function coefficients are used as a multiplier when the variables have been standardised to a mean of zero and a variance of one. For standardised canonical discriminant, the weighted average of 1.042 separates the observations into the two groups for Unemployed, while the weighted average of 0.031 into the two groups for Literacy and weighted average of 0.222 into the two groups for State. For centroids function coefficients, the weighted average of 0.978 separates the observations into the two groups for Unemployed, while the weighted average of 0.093 into the two groups for Literacy and weighted average of 0.052 into the two groups for State.

Table 8.Classification Statistics, percentages and Prior Probabilities for Groups

Group		Predicted Group Membership		Total	Prior	Functions at Group Centroids
		Low	High			
Count	Low	18	0	18	0.500	-1.601
	High	1	18	19	0.500	1.517
%	Low	100.0	.0	100.0	1.000	
	High	5.3	94.7	100.0		

The centroids in Table 8 show the mean values for the discriminant score for each of the groups. Low has a negative value of -1.601 because a wrongly classified state from a high rate of unemployment is included in the group, while High has a positive value of 1.517. All the states with a low rate of unemployment are correctly classified and 94.7% of the states with high rate unemployment were correctly classified and 5.3% were wrongly classified. It was assumed that the prior probabilities for the group are equal as set by default of the SPSS package version 23 that was used.

The casewise statistics on Table 9 (*end of the chapters*) shows that only one state was misclassified, which have the smallest Squared Mahalanobis Distance to Centroid, and then showed misrepresentation between actual and predicted Group. The state is coded 5 which is Bauchi state. The unemployment rate was low and not high.

Conclusion

The analysis shows that 18 states were correctly classified. The classification code are: 1, 3, 6, 8, 9, 10, 12, 15, 16, 17, 18, 19, 25, 31, 32, 33, 35, and 37 were correct and that these state rated high level of unemployment. These codes stand for Abia, Akwa-Ibom, Bayelsa, Borno, Cross Rivers, Delta, Edo, Gombe, Imo, Jigawa, Kaduna, Kano, Nasarawa, Plateau, Rivers, Sokoto, Yobe, and FCT Abuja. However, 18 states with codes 2, 4, 7, 11, 13, 14, 20, 21, 22, 23, 24, 26, 27, 28, 29, 30, 34, and 36 are rated accurately as states with low unemployment. The are Adamawa, Anambra, Benue, Ebonyi, Ekiti, Enugu, Katsina, Kebbi, Kogi, Kwara, Lagos, Niger, Ogun, Ondo, Osun, Oyo, Taraba, and Zamfara. Code number 5, which stands for Bauchi was misclassified. The respective p-values and discriminant Score function are indicated in Table 9. The serial coding followed the alphabetical order of the arrangement of States in Nigeria.

References

1. Alayande, S. A. and Bashiru K. A., (2015). An Overview and Application of Discriminant Analysis in Data Analysis. Journal of Mathematics Vol (11), Issue 1, PP 12-15 www.iosrjournals.org.
2. Grimm, L. G. and Yarnold, P. R.(1995): Reading and Understanding Multivariate Statistics. Washington, D.C.: *American Psychological Association*.
3. Huberty, C. J. and Olejnik, S. (2006): Applied MANOVA and Discriminant Analysis, Second Edition. Hoboken, *New Jersey: John Wiley and Sons, Inc*.
4. Ramayah, T. Hazlima, N. A. Hasliza, A. H. Siti, R. Z.M and Lo, M (2010): discriminant Analysis: An Illustrated example. *African Journal of Business Management* 4(9).
5. Stevens, J. P. (2002): Applied Multivariate Statistics for the Social Sciences, Fourth Edition. Mahwah, *New Jersey: Lawrence Erlbaum Associates*, Inc.
6. Tatsuoka, M. M., (1971): Multivariate Analysis: Techniques for Educational and Psychological Research.*New York: John Wiley and Sons.*
7. Unemployment rate in Nigeria by state. Third Quarters (2018) and Adult Literacy Data, Nigeria. National Bureau of Statistics: www.nigerianstat.gov.ng, accessed 12th June, 2019.

International Journal of Transformation in Applied Mathematics & Statistics
Vol. 3, Issue 1 – 2020
ISSN: 2581-7620

Appendices A: Spss Output

Analysis Case Processing Summary			
Unweighted Cases		N	Percent
Valid		37	59.7
Excluded	Missing or out-of-range group codes	0	.0
	At least one missing discriminating variable	0	.0
	Both missing or out-of-range group codes and at least one missing discriminating variable	25	40.3
	Total	25	40.3
Total		62	100.0

Group Statistics					
GROUP		**Mean**	**Std. Deviation**	**Valid N (listwise)**	
				Unweighted	**Weighted**
LOW	UNEMPLOY	17.6167	3.55681	18	18.000
	LITERACY	68.3167	11.41378	18	18.000
	STATE	20.6111	10.05946	18	18.000
HIGH	UNEMPLOY	28.9842	3.88219	19	19.000
	LITERACY	70.0263	9.85003	19	19.000
	STATE	17.4737	11.56320	19	19.000
Total	UNEMPLOY	23.4541	6.83287	37	37.000
	LITERACY	69.1946	10.52523	37	37.000
	STATE	19.0000	10.82436	37	37.000

Minitab, output.

Paired T-Test and CI: low, high

Paired T for low - high

N Mean StDev SE Mean

low 3 35.5 28.4 16.4

high 3 38.8 27.6 15.9

Difference 3 -3.31 7.38 4.26

95% CI for mean difference: (-21.66, 15.03)

T-Test of mean difference = 0 (vs ≠ 0): T-Value = -0.78 P-Value = 0.518

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
UNEMPLOY	.289	85.956	1	35	.000
LITERACY	.993	.239	1	35	.628
STATE	.978	.772	1	35	.386

Analysis 1

Box's Test of Equality of Covariance Matrices

Log Determinants		
GROUP	Rank	Log Determinant
LOW	3	11.672
HIGH	3	11.847
Pooled within-groups	3	11.998
The ranks and natural logarithms of determinants printed are those of the group covariance matrices.		

Test Results		
Box's M		8.263
F	Approx.	1.248
	df1	6
	df2	8798.394
	Sig.	.278
Tests null hypothesis of equal population covariance matrices.		

Summary of Canonical Discriminant Functions

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	2.568[a]	100.0	100.0	.848
a. First 1 canonical discriminant functions were used in the analysis.				

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	Df	Sig.
1	.280	42.611	3	.000

Standardized Canonical Discriminant Function Coefficients	
	Function
	1
UNEMPLOY	1.042
LITERACY	.031
STATE	.222

Structure Matrix	
	Function
	1
UNEMPLOY	.978
STATE	-.093
LITERACY	.052
Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions Variables ordered by absolute size of correlation within function.	

Functions at Group Centroids	
GROUP	Function
	1
LOW	-1.601
HIGH	1.517
Unstandardized canonical discriminant functions evaluated at group means	

Classification Statistics

Classification Processing Summary		
Processed		62
Excluded	Missing or out-of-range group codes	0
	At least one missing discriminating variable	25
Used in Output		37

Prior Probabilities for Groups			
GROUP	Prior	Cases Used in Analysis	
		Unweighted	Weighted
LOW	.500	18	18.000
HIGH	.500	19	19.000
Total	1.000	37	37.000

Classification Function Coefficients		
	GROUP	
	LOW	HIGH
UNEMPLOY	1.549	2.421
LITERACY	.664	.673
STATE	.462	.526
(Constant)	-41.768	-63.925
Fisher's linear discriminant functions		

Case		Actual	Highest Group					Second Highest Group			Discriminant Scores
Number		Group	Predicted Group	P(D>d | G=g) P	df	P(G=g | D=d)	Squared Mahalanobis Distance to Centroid	Group	P(G=g | D=d)	Squared Mahalanobis Distance to Centroid	Function 1
Original	1	2	2	0.67	1	0.998	0.182	1	0.002	12.566	1.944
	2	1	1	0.601	1	0.962	0.274	2	0.038	6.735	-1.078
	3	2	2	0.03	1	1	4.699	1	0	27.942	3.685
	4	1	1	0.722	1	0.997	0.127	2	0.003	12.072	-1.958
	5	2	1**	0.188	1	0.68	1.735	2	0.32	3.244	-0.284
	6	2	2	0.425	1	0.999	0.637	1	0.001	15.338	2.315
	7	1	1	0.667	1	0.971	0.185	2	0.029	7.227	-1.171
	8	2	2	0.654	1	0.998	0.2	1	0.002	12.716	1.965
	9	2	2	0.765	1	0.997	0.089	1	0.003	11.676	1.816
	10	2	2	0.25	1	0.781	1.325	1	0.219	3.869	0.366
	11	1	1	0.434	1	0.919	0.611	2	0.081	5.459	-0.82
	12	2	2	0.224	1	0.745	1.477	1	0.255	3.62	0.301
	13	1	1	0.564	1	0.955	0.334	2	0.045	6.455	-1.024
	14	1	1	0.875	1	0.988	0.025	2	0.012	8.765	-1.444
	15	2	2	0.533	1	0.949	0.388	1	0.051	6.225	0.894
	16	2	2	0.831	1	0.985	0.045	1	0.015	8.439	1.304
	17	2	2	0.489	1	0.937	0.479	1	0.063	5.886	0.825
	18	2	2	0.553	1	0.953	0.351	1	0.047	6.378	0.924
	19	2	2	0.49	1	0.999	0.477	1	0.001	14.506	2.207
	20	1	1	0.325	1	1	0.967	2	0	16.824	-2.585
	21	1	1	0.503	1	0.941	0.448	2	0.059	5.996	-0.932
	22	1	1	0.533	1	0.949	0.388	2	0.051	6.226	-0.978

	23	1	1	0.343	1	0.871	0.898	2	0.129	4.713	-0.654
	24	1	1	0.473	1	0.999	0.514	2	0.001	14.71	-2.318
	25	2	2	0.737	1	0.978	0.112	1	0.022	7.745	1.182
	26	1	1	0.33	1	0.861	0.95	2	0.139	4.594	-0.626
	27	1	1	0.856	1	0.996	0.033	2	0.004	10.89	-1.783
	28	1	1	0.432	1	0.999	0.618	2	0.001	15.244	-2.387
	29	1	1	0.054	1	1	3.703	2	0	25.427	-3.526
	30	1	1	0.065	1	1	3.401	2	0	24.625	-3.445
	31	2	2	0.624	1	0.998	0.241	1	0.002	13.022	2.007
	32	2	2	0.016	1	1	5.785	1	0	30.508	3.922
	33	2	2	0.621	1	0.965	0.245	1	0.035	6.884	1.022
	34	1	1	0.519	1	0.945	0.415	2	0.055	6.119	-0.957
	35	2	2	0.763	1	0.997	0.091	1	0.003	11.696	1.819
	36	1	1	0.643	1	0.968	0.215	2	0.032	7.045	-1.137
	37	2	2	0.364	1	0.884	0.824	1	0.116	4.886	0.609

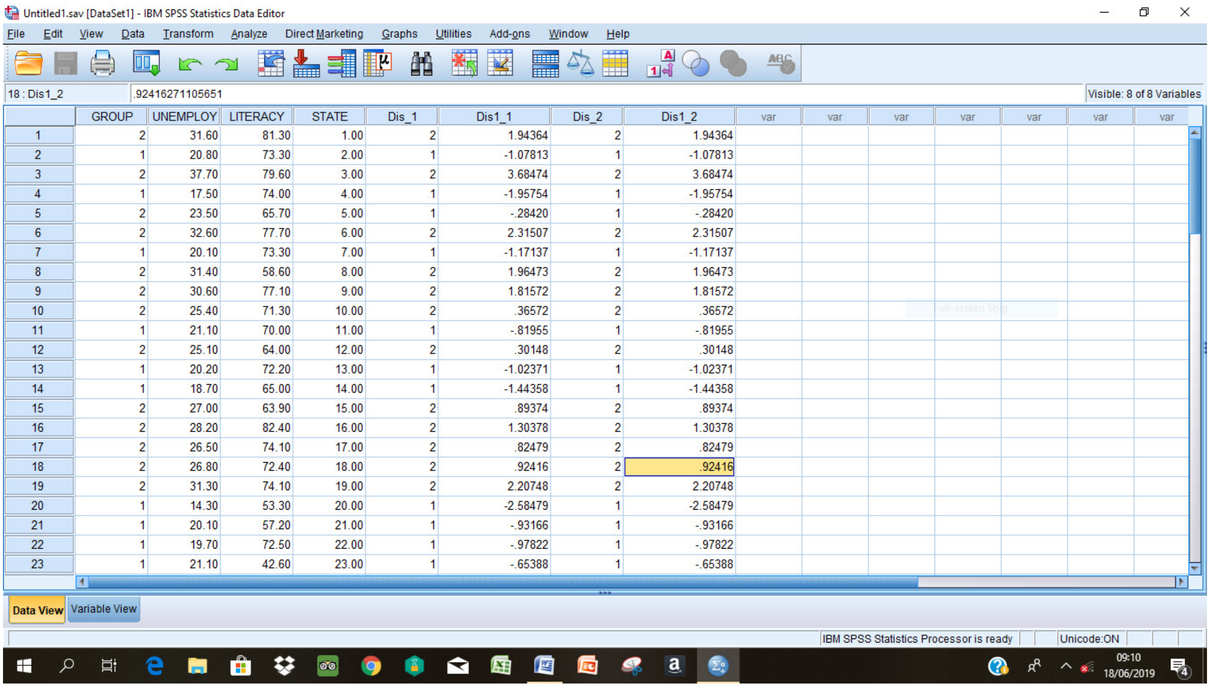
Table 14.Casewise Statistics

Classification Results[a]					
		GROUP	Predicted Group Membership		**Total**
			LOW	HIGH	
Original	Count	LOW	18	0	18
		HIGH	1	18	19
	%	LOW	100.0	.0	100.0
		HIGH	5.3	94.7	100.0
a. 97.3% of original grouped cases correctly classified.					

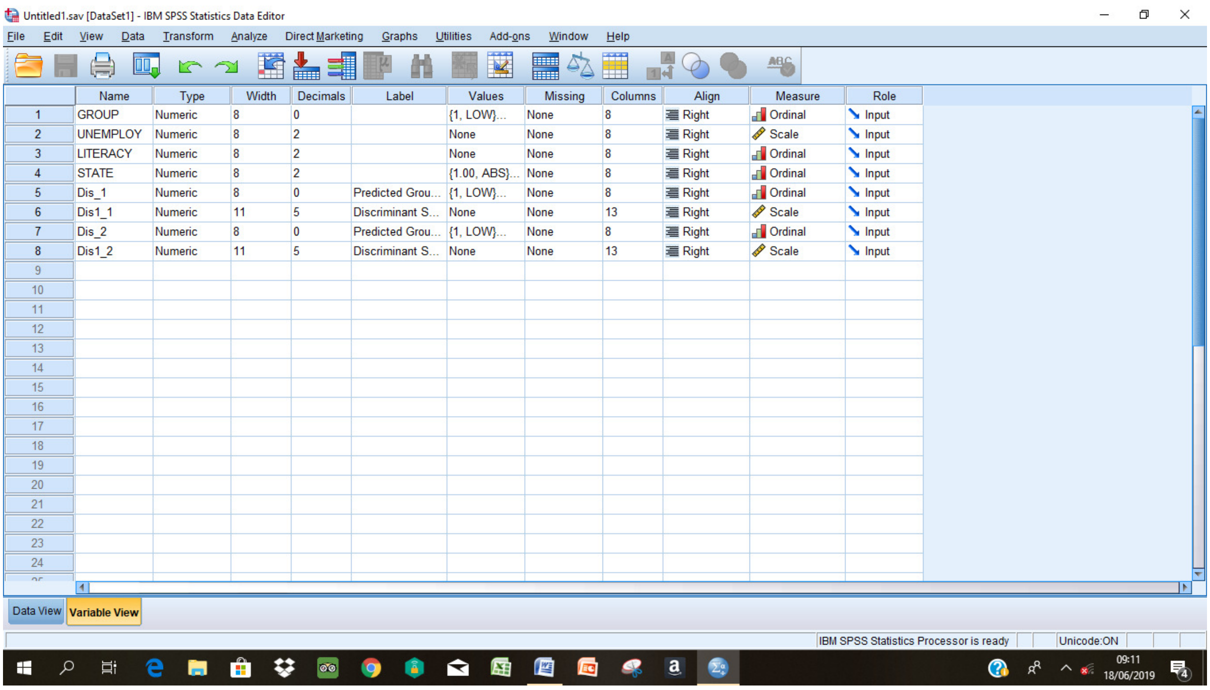
International Journal of Transformation in Applied Mathematics & Statistics
Vol. 3, Issue 1 – 2020
ISSN: 2581-7620

Appendix

Data sets on the SPSS spreadsheet



	GROUP	UNEMPLOY	LITERACY	STATE	Dis_1	Dis1_1	Dis_2	Dis1_2
1	2	31.60	81.30	1.00	2	1.94364	2	1.94364
2	1	20.80	73.30	2.00	1	-1.07813	1	-1.07813
3	2	37.70	79.60	3.00	2	3.68474	2	3.68474
4	1	17.50	74.00	4.00	1	-1.95754	1	-1.95754
5	2	23.50	65.70	5.00	1	-.28420	1	-.28420
6	2	32.60	77.70	6.00	2	2.31507	2	2.31507
7	1	20.10	73.30	7.00	1	-1.17137	1	-1.17137
8	2	31.40	58.60	8.00	2	1.96473	2	1.96473
9	2	30.60	77.10	9.00	2	1.81572	2	1.81572
10	2	25.40	71.30	10.00	2	.36572	2	.36572
11	1	21.10	70.00	11.00	1	-.81955	1	-.81955
12	2	25.10	64.00	12.00	2	.30148	2	.30148
13	1	20.20	72.20	13.00	1	-1.02371	1	-1.02371
14	1	18.70	65.00	14.00	1	-1.44358	1	-1.44358
15	2	27.00	63.90	15.00	2	.89374	2	.89374
16	2	28.20	82.40	16.00	2	1.30378	2	1.30378
17	2	26.50	74.10	17.00	2	.82479	2	.82479
18	2	26.80	72.40	18.00	2	.92416	2	.92416
19	2	31.30	74.10	19.00	2	2.20748	2	2.20748
20	1	14.30	53.30	20.00	1	-2.58479	1	-2.58479
21	1	20.10	57.20	21.00	1	-.93166	1	-.93166
22	1	19.70	72.50	22.00	1	-.97822	1	-.97822
23	1	21.10	42.60	23.00	1	-.65388	1	-.65388



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	GROUP	Numeric	8	0		{1, LOW}...	None	8	Right	Ordinal	Input
2	UNEMPLOY	Numeric	8	2		None	None	8	Right	Scale	Input
3	LITERACY	Numeric	8	2		None	None	8	Right	Ordinal	Input
4	STATE	Numeric	8	2		{1.00, ABS}...	None	8	Right	Ordinal	Input
5	Dis_1	Numeric	8	0	Predicted Grou...	{1, LOW}...	None	8	Right	Ordinal	Input
6	Dis1_1	Numeric	11	5	Discriminant S...	None	None	13	Right	Scale	Input
7	Dis_2	Numeric	8	0	Predicted Grou...	{1, LOW}...	None	8	Right	Ordinal	Input
8	Dis1_2	Numeric	11	5	Discriminant S...	None	None	13	Right	Scale	Input

Table 9.Casewise Statistics (Classification)

Case Number		Actual Group	Highest Group					Second Highest Group			Discriminant Scores
			Predicted Group	P(D>d | G=g)		P (G=g | D=d)	Squared Mahalanobis Distance to Centroid	Group	P (G=g | D=d)	Squared Mahalanobis Distance to Centroid	Function 1
				P	Df						
Original	1	2	2	0.670	1	0.998	0.182	1	0.002	12.566	1.944
	2	1	1	0.601	1	0.962	0.274	2	0.038	6.735	-1.078
	3	2	2	0.030	1	1.000	4.699	1	0.000	27.942	3.685
	4	1	1	0.722	1	0.997	0.127	2	0.003	12.072	-1.958
	5	**2**	**1****[**]**	**0.188**	**1**	**0.680**	**1.735**	**2**	**0.320**	**3.244**	**-0.284**
	6	2	2	0.425	1	0.999	0.637	1	0.001	15.338	2.315
	7	1	1	0.667	1	0.971	0.185	2	0.029	7.227	-1.171
	8	2	2	0.654	1	0.998	0.200	1	0.002	12.716	1.965
	9	2	2	0.765	1	0.997	0.089	1	0.003	11.676	1.816
	10	2	2	0.250	1	0.781	1.325	1	0.219	3.869	0.366
	11	1	1	0.434	1	0.919	0.611	2	0.081	5.459	-0.820
	12	2	2	0.224	1	0.745	1.477	1	0.255	3.620	0.301
	13	1	1	0.564	1	0.955	0.334	2	0.045	6.455	-1.024
	14	1	1	0.875	1	0.988	0.025	2	0.012	8.765	-1.444
	15	2	2	0.533	1	0.949	0.388	1	0.051	6.225	0.894
	16	2	2	0.831	1	0.985	0.045	1	0.015	8.439	1.304
	17	2	2	0.489	1	0.937	0.479	1	0.063	5.886	0.825
	18	2	2	0.553	1	0.953	0.351	1	0.047	6.378	0.924
	19	2	2	0.490	1	0.999	0.477	1	0.001	14.506	2.207
	20	1	1	0.325	1	1.000	0.967	2	0.000	16.824	-2.585
	21	1	1	0.503	1	0.941	0.448	2	0.059	5.996	-0.932

	22	1	1	0.533	1	0.949	0.388	2	0.051	6.226	-0.978
	23	1	1	0.343	1	0.871	0.898	2	0.129	4.713	-0.654
	24	1	1	0.473	1	0.999	0.514	2	0.001	14.710	-2.318
	25	2	2	0.737	1	0.978	0.112	1	0.022	7.745	1.182
	26	1	1	0.330	1	0.861	0.950	2	0.139	4.594	-0.626
	27	1	1	0.856	1	0.996	0.033	2	0.004	10.890	-1.783
	28	1	1	0.432	1	0.999	0.618	2	0.001	15.244	-2.387
	29	1	1	0.054	1	1.000	3.703	2	0.000	25.427	-3.526
	30	1	1	0.065	1	1.000	3.401	2	0.000	24.625	-3.445
	31	2	2	0.624	1	0.998	0.241	1	0.002	13.022	2.007
	32	2	2	0.016	1	1.000	5.785	1	0.000	30.508	3.922
	33	2	2	0.621	1	0.965	0.245	1	0.035	6.884	1.022
	34	1	1	0.519	1	0.945	0.415	2	0.055	6.119	-0.957
	35	2	2	0.763	1	0.997	0.091	1	0.003	11.696	1.819
	36	1	1	0.643	1	0.968	0.215	2	0.032	7.045	-1.137
	37	2	2	0.364	1	0.884	0.824	1	0.116	4.886	0.609
Footnote: **. Misclassified case; code 1= low and code 2= high											