

Discriminant Analysis: An Analysis of Its Predictship Function

Ogbogo Stella

Department of Educational Psychology, Guidance & Counselling, Faculty of Education, University of Port Harcourt, Rivers State, Nigeria

Abstract

Discriminant analysis as a multivariate statistical method was reviewed in this paper. Alongside this was the general goal of discriminate analysis which is to predict membership from a set of predictor as well as to classify individuals into one of two or more alternative groups on the basis of a set of measurements. Discriminant analysis builds a predictive model for group membership. This model is composed of a discriminant function (or, for more than two groups, a set of discriminant functions). The discriminate function coefficients gives the contribution of each variable to the function. In order to derive more substantive "meaningful" labels for the discriminant functions, one can also examine the factor structure matrix with the correlations between the variables and the discriminant functions as well as other statistics associated with discriminate analysis.

Keywords: Discriminate analysis, discriminate function, prediction, classification discriminate function coefficients.

DOI: 10.7176/JEP/10-5-04

Introduction

Discriminant analysis (DA) is a multivariate analysis. It is a method that is used to find the linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier or, more commonly, for dimension reduction before later classification. Discriminant analysis is a technique that is used in research to analyze the research data when the criterion or the dependent variable is categorical and the predictor or the independent variable is interval in nature. The term categorical variable means that the dependent variable is divided into a number of categories (Tabachnick & Fidell 2013). For example, three types of anxiety, Social anxiety A, School anxiety B and Class anxiety C can be the categorical dependent variable. In simple terms, discriminant function analysis is classification - the process or act of distributing things of the same type into groups, classes or categories. Discriminant analysis is a statistical method that is used in research to help understand the relationship between a "dependent variable" and one or more "independent variables." A dependent variable is the variable that a researcher is trying to explain or predict from the values of the independent variables (Lawler, 2018).

Discriminant analysis builds a predictive model for group membership. The model is composed of a discriminant function (or, for more than two groups, a set of discriminant functions) based on linear combinations of the predictor variables that provide the best discrimination between the groups. The functions are generated from a sample of cases for which group membership is known; the functions can then be applied to new cases that have measurements for the predictor variables but have unknown group membership. Discriminant or discriminant function analysis is a parametric technique to determine which weightings of quantitative variables or predictors best discriminate between 2 or more than 2 groups of cases and do so better than chance Cramer, 2003 in Ramayah et al.

General Purpose

The purpose or goal of discriminate analysis is to predict group membership from a set of predictors. It is used to classify individuals into one of two or more alternative groups on the basis of a set of measurements. (Li,2012). Discriminant function analysis is used to determine which variables discriminate between two or more naturally occurring groups. Discriminant analysis is most often used to help a researcher predict the group or category to which a subject belongs (Hill & Lewicki 2007). In research, it helps to examine whether significant differences exist among the groups, in terms of the predictor variables. It also evaluates the accuracy of the classification. As in statistics, everything is assumed up until infinity, so in a case, when the dependent variable has two categories, then the type used is two-group discriminant analysis. That is a complicated case, in which the dependent variable has more than two categories. If the dependent variable has three or more than three categories, then the type used is multiple discriminant analysis. For example, anxiety might have been divided into three groups: Social anxiety, school anxiety and class anxiety. Discriminant analysis allows for such a case, as well as many more categories. The major distinction to the types of discriminant analysis is that for a two group, it is possible to derive only one discriminant function. On the other hand, in the case of multiple discriminant analysis, more than one discriminant function can be computed (Lawler 2017). For instance, an educational researcher may want to investigate which variables discriminate between students who have anxiety for (1) Social setting, (2) school, or (3) class. For that purpose the researcher could collect data on numerous variables on the students' and

see which students will naturally fall into one of the three categories. Another instance can be when a medical researcher may record different variables relating to patients' backgrounds in order to learn which variables best predict whether a patient is likely to recover completely (group 1), partially (group 2), or not at all (group 3). In addition to the examples given below, DA is applied in so many areas like bankruptcy prediction, face recognition, marketing, science, so on.

History, Similarities/Difference with Other Procedures

The original dichotomous discriminant analysis was developed by Sir Ronald Fisher in 1936. It is different from an ANOVA or MANOVA, which is used to predict one (ANOVA) or multiple (MANOVA) continuous dependent variables by one or more independent categorical variables. Discriminate analysis has had its earliest and most widespread educational research applications in the areas of vocational and career development. (Lohnes,1988). Because education prepares people for a variety of positions in the occupational structure prevalent in their societies, this thus became an important class of educational research that was concerned with the testing of theories about the causes of occupational placement and or the estimation of the prediction equation for allocating positions or anticipating such allocations (Johnes, 1988).

Discriminate Analysis is closely related to Analysis of Variance (ANOVA), Cluster Analysis, MANOVA, Principal component analysis and factor analysis, Logistic Regression, MANOVA and regression analysis, which also attempt to express one dependent variable as a linear combination of other features or measurements. Though similar to these procedures, there exists as well, differences between discriminate analysis and these other procedures.

Regression analysis according to Iweka, (2017), refers to the statistical procedures for which the primary purpose is the estimation of scores on one variable from scores on other variables. Regression analysis is widely used for prediction and forecasting. In many ways, discriminant analysis parallels multiple regression analysis. The main difference between these two techniques is that regression analysis deals with a continuous dependent variable, while discriminant analysis must have a discrete dependent variable. Discriminate Analysis is closely related to Analysis of variance (ANOVA), However, ANOVA uses categorical independent variable and a continuous dependent variable whereas discriminant analysis has continuous independent variables and a categorical dependent variable (*i.e.* the class label).

Logistic regression and probit regression are more similar to Discriminate Analysis than ANOVA is, as they also explain a categorical variable by the values of continuous independent variables. These other methods are preferable in applications where it is not reasonable to assume that the independent variables are normally distributed, which is a fundamental assumption of the DA method. Discriminant function analysis is very similar to Logistic regression, and both can be used to answer the same research questions. Logistic regression does not have as many assumptions and restrictions as discriminant analysis. However, when discriminant analysis' assumptions are met, it is more powerful than logistic regression. It is often preferred to discriminate analysis as it is more flexible in its assumptions and types of data that can be analyzed (Pousen & French 2018). Logistic regression can handle both categorical and continuous variables, and the predictors do not have to be normally distributed, linearly related, or of equal variance within each group (Tabachnick and Fidell 2013).

Discriminant analysis is closely related also to Cluster analysis. In DA, the group information is known. Or there is a set of samples with both known independent variables (*i.e.* group information) and dependent variables. While, in CA, only the information of the independent variables is known. All varieties of discriminant analysis require prior knowledge of the classes, usually in the form of a sample from each class. In cluster analysis, the data do not include information about class membership; the purpose is to construct a classification.

Canonical correlation analysis is a general multivariate technique in which discriminate analysis is a special case of it. It is used to analyze the relationship that exists between two sets of variables. These sets of variables have variables under them that have been combined to form a set. These sets of variables are then combined to form what is called a variate. Some forms of canonical correlation is imbedded in the discriminate analysis process but DA analysis goes a bit further. The goal of discriminate analysis is not just to find if a set of variables relates to or predict a set of another variable but it aims to discriminate and classify as well.

DA is also closely related to principal component analysis (PCA) and factor analysis though there exists as well.

The mathematics of discriminant analysis is related very closely to MANOVA. In fact, the roles of the variables are simply reversed. The classification (factor) variable in the MANOVA becomes the dependent variable in discriminant analysis. The dependent variables in the MANOVA become the independent variables in the discriminant analysis.

Assumptions

The following assumptions guide Discriminate Analysis

Sample size: Unequal sample sizes are acceptable. The sample size of the smallest group needs to exceed the

number of predictor variables. As a “rule of thumb”, the smallest sample size should be at least 20 for a few (4 or 5) predictors. The maximum number of independent variables is $n - 2$, where n is the sample size. While this low sample size may work, it is not encouraged, and generally it is best to have 4 or 5 times as many observations and independent variables. (Poulsen & French 2018)

Normal distribution: It is assumed that the data (for the variables) represent a sample from a multivariate normal distribution. Predictor variables should have a multivariate normal distribution. One can examine whether or not variables are normally distributed with histograms of frequency distributions, also in addition to this, one may use specific tests for normality.

Homogeneity of variances/covariances. It is assumed that the variance/covariance matrices of variables are homogeneous across groups. That is within-group variance-covariance matrices should be equal across groups

Correlations between means and variances. The major "real" threat to the validity of significance tests occurs when the means for variables across groups are correlated with the variances (or standard deviations). This thus should be avoided

Non-multicollinearity: If one of the independent variables is very highly correlated with another, or one is a function (e.g., the sum) of other independents, then the tolerance value for that variable will approach 0 and the matrix will not have a unique discriminant solution. There must also be low multicollinearity of the independents.

Outliers and Missing data: DA is highly sensitive to the inclusion of outliers and cases where data are missing. To avoid violating this assumption, one should run a test for univariate and multivariate outliers for each group, and transform or eliminate them.

Linearity; Discriminant analysis assumes linear relations among the independent variables.. Therefore, the occurrence of a curvilinear relationship reduces the power and the discriminating ability of the discriminant equation. Predictive power can decrease with an increased correlation between predictor variables.

Independence Membership in a group is assumed to be mutually exclusive (that is, no case belongs to more than one group) and collectively exhaustive (that is, all cases are members of a group).

Steps/Approach/Computation

As the mathematical methods used in discriminant analysis are complex, they are described here only in general terms, however the fundamental equations are described here

Fundamental Equations for Discriminate Analysis

First, variance in the set of predictors is partitioned into two sources: variance attributable to differences between groups and variance attributable to differences within groups. After which cross-products matrices are formed.

$$S_{total} = S_{bg} + S_{wg}$$

The total cross-products matrix (S_{total}) is partitioned into a cross-products matrix associated with differences between groups (S_{bg}) and a cross—products matrix of differences within groups (S_{wg}). This is akin to the equation for MANOVA, which follows as well the equation of ANOVA where variance is partitioned into sums of squares sums of square between and within.

The determinants are calculated for these matrices and used to calculate a test statistic – i.e Mahalanobis’s D^2 , Wilks Lambda, Rao’s r , Roys characteristics, Hotelling’s trace, Pillia’s Criterion are used for test of significance. (they are in themselves descriptive statistics though inferential statistics is applied to them). In this case, Wilks Lambda is presented.

Wilks Lambda follows the equation:

$$\hat{\Lambda} = \frac{S_{wg}}{S_{bg} + S_{wg}}$$

Tables for calculating Wilks lambda directly are rare however, an approximation to F has been derived that closely fits Wilks Lambda ($\hat{\Lambda}$). The next procedure for calculating approximate F is based on Wilks is lambda Next is the approximate F ratio

$$F \text{ approximate } cdf_1 df_2 = \left(\frac{1 - y}{y} \right) \left(\frac{df_2}{df_1} \right)$$

For cases where n is equal in all groups:

$$Y = \hat{\Lambda}^{1/p} \quad p = \# \text{ of predictor variables}$$

$$s = \sqrt{\frac{p^2 (df_{effect})^2 - 4}{p^2 + (df_{effect})^2 - 5}}$$

$df_{error} = \text{number of groups times } (n-1): k(n-1)$
 $Df_1 = p(df_{effect})$

$$Df_2 = s \left[(df_{error}) - \frac{p - df_{effect}}{2} \right] - \left[\frac{p(df_{effect}) - 2}{2} \right]$$

$Df_{effect} = \text{number of groups minus one } (k - 1)$

For unequal n between groups, this is modified only by changing the df_{error} , to equal the number of data points in all groups minus the number of groups ($N - k$).

If F is significantly then one can say that a set of predictors can be used to classify or distinguish a group of category. This is a test of overall relationship between groups predictors i.e groups can be distinguished on the basis of the predictor variables if an overall relationship is found, the next thus is to determine or examine the discriminant functions that compose the overall relationship.

The discriminate function refers to a linear combination of a set of variables. Discriminate functions are also known toots, canonical variables, principal components, dimensions etc, depending on the statistical technique (Tabachnick & Fidell 2013). A discriminate function is composed of sum of the set of predictors each weighted by a coefficient. The maximum number of discriminant functions is either (1) the number of predictors or (2) the degrees of freedom fir groups, whichever is smaller. For Instance when there are three groups (and four predictors), there are potentially two discriminate functions contributing to the overall relationship. And, when the overall relationship is statistically significant, at least the first discriminant function is very likely to be significant, and both may be significant.

The linear combination for a discriminant analysis, also known as the discriminant function, is derived from an equation that takes the following form:

$$Z_{ik} = b_{0i} + b_{1i} X_{1k} + b_{ji} X_{jk}$$

Where:

Z_{jk} discriminant score of discriminant function i for object k .

$i = 1, \dots, G - 1$

X_{jk} independent variable j for object $k, j = 1, 2, \dots, J$

b_{ji} discriminant weight (coefficient) for independent variable j and discriminant function i

b_{0i} constant of discriminant function i

Or this equation as well solves for the (standardized) discriminant function score for the i th function,

$$D_j = D_{11z1} + D_{12z2} + \dots + d_{ip}z_p$$

An individual's standardized score on the i th discriminant function (D_i) is found by multiplying the standardized score on each predictor (z) by its standardized discriminant function coefficient (d_i) and then adding the products for all predictors.

Discriminant function coefficients are found in the same manner as are coefficients for regression or canonical variates. In fact, canonical correlation is an aspect of Discriminate analysis with group membership on one side of the equation and predictors on the other, where successive canonical variates (here called discriminant functions) are computed. (Tabachnick & Fidell 2013). Canonical coefficients are the elements of some eigenvectors (these eigenvectors, are V , of $S_w^{-1}S_A$). A discriminant function score for a case, then, can also be produced by multiplying the raw score on each predictor by its associated unstandardized discriminant function coefficient adding the products over all predictors, and adding a constant to adjust for the means. The mean of each discriminant function over all cases is zero, because the mean of each predictor, when standardized, is zero. The standard deviation of each D_i is 1. (Tabachnick & Fidell 1996).

Just as D_i can be calculated for each case, a mean value of D_i can be calculated for each group. The members of each group considered together have a mean score of a discriminant function which is the distance of the group in standard deviation units, from the zero mean of the discriminant function. Group means on D_i are typically called centroids. Discriminate function score is computed for each case, thereafter discriminate function coefficient mean is computed for each group. The discriminate function coefficient mean for each group as earlier mentioned is called a centroid. A large centroid between two groups means that the discriminate function separates the groups. The discriminate coefficient reveals as well it is predictive power to the discriminate Function. When the sign is ignored, each weight (coefficient) represents the relative contribution of its associated variable to that function. Independent variables with relatively larger weights contribute more to the discriminating power of the function than do variables with smaller weights. The discriminant function coefficients are analogous regression coefficients and they range between values of -1.0 and 1.0. The magnitudes

of the coefficients also tell us something about the relative contributions of the independent variables. The closer the value of a coefficient is to zero, the weaker it is as a predictor of the dependent variable. On the other hand, the closer the value of a coefficient is to either 1.0 or -1.0, the stronger it is as a predictor of the dependent variable. (Lawler 2017)

For each discriminate function, there is a discriminant loadings (structure correlations). This measure the simple linear correlation between each independent variable and the discriminant function. The discriminant loadings reflect the variance that the independent variables share with the DA function and can be interpreted as assessing the relative contribution of each independent variable to the DA function. For each Discriminant Function as well, a canonical correlation is found for each discriminant function. Canonical correlations are found by solving for the eigenvalues and eigenvectors of a correlation matrix. Eigenvalue is a form of squared correlation coefficients, represents overlapping variance among variables, in this case between predictors and groups. Other vectors computed for in Discriminate Function is within-group, between group and overall covariance matrix. These takes the form of the equations below:

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

$$rcj = \sqrt{\frac{\lambda_j}{1 + \lambda_j}}$$

Various other matrices are often considered during a discriminant analysis. The overall covariance matrix, T , is given by:

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

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

$$W = \left(\frac{1}{N - K} \right) S_W$$

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

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

The standardized canonical coefficients are given by:

$$v_{ij} \sqrt{w_{ij}}$$

Where v_{ij} are the elements of V and w_{ij} are the elements of W .

The correlations between the independent variables and the canonical variates are given by:

$$\text{Corr}_{jk} = \frac{1}{\sqrt{w_{ij}}} \sum_{i=1}^p v_{ik} w_{ji}$$

After all possible discriminant function is obtained, successive discriminant functions are evaluated for significance. Each function with its given discriminant score is evaluated to determine how well it predicts group placement. That is evaluation of each Discriminate function if they are significant or not. Using Wilks Lambda, (which is transformed to an approximate F), the discriminatory power of each significant function is evaluated. A large F usually indicates a significantly larger discriminatory/predictive power. Although many a time, the first two discriminate function usually accounts for a lion share of the discriminate power. Discriminant analysis then, finds “good” regions of to minimize classification error, therefore leading to a high percent correct classified in the classification table. (Hardle, & Simar, 2007).

If there are only two groups, discriminant function scores can be used to classify cases into groups. A case is classified into one group if its D_1 score is above zero, and into the other group if the D_i score is below zero. With numerous groups, classification is possible from the discriminant function, but it is simpler to use the procedure in the following section.

To assign cases into groups, a classification equation is developed for each group. For instance, when there are three groups, three classification equations are developed, for the three groups. Data for case are inserted into each classification equation to develop a classification score for each group for the case. The case is assigned to the group for which it has the highest classification score.

In its simplest form, the basic classification equation for the j^{th} group ($j = 1, 2, \dots, k$) is

$$C_j = c_{j0} + c_{j1}X_1 + C_{j2}X_2 + \dots + c_{jp} X_p$$

A score on the classification function for group j (C_j) is found by multiplying the raw score on each predictor (X) by its associated classification function coefficient (c_j) summing over all predictors, and adding a constant c_{j0} .

Classification coefficients, c are found from the means of the p predictors and the pooled within-group variance—covariance matrix, W . The within-group covariance matrix is produced by dividing each element in the cross-product matrix, S_{wg} by the within-group degrees of freedom, $N-k$. In matrix form,

$$C_j = W^{-1}M_j$$

The column matrix of classification coefficients for group j ($C_j = c_{j1}, c_{j2}, \dots, c_{jp}$) is found by multiplying the inverse of the within-group variance—covariance matrix W^{-1} , by a column matrix of means for group j on the p variables ($M_j = X_{j1}, X_{j2}, \dots, X_{jp}$).

The constant for group j , c_{j0} , is found as follows:

$$C_{j0} = \left(\frac{1}{2}\right) C_j M_j$$

The constant for the classification function for group j (r) is formed by multiplying $\frac{1}{2}$ times the transpose of the column matrix of classification coefficient's for group j (C_j) times the column matrix of means for group j (M_j).

Multiplying W^{-1} by the column matrix of means for the first group gives the matrix of classification coefficients for that group,

This simple classification scheme is most appropriate when equal group sizes are expected in the population. If unequal group sizes are expected, the classification procedure can be modified by setting a priori probabilities to group size. The classification equation for group j (C_j) then becomes.

$$C_j = c_{j0} + \sum_{i=1}^p c_{ji} X_i + \ln(n_j / N)$$

Where n_j is the size of group j and N is the total sample size.

The number of equations are usually based on the number of groups. Thus data for each case is inserted into each group classification equation to have a classification score. The score is assigned into the group for which it has the highest score.

Fundamentally, the following statistics apply to the computation of Discriminate Analysis

Statistics. For each variable: means, standard deviations, univariate ANOVA. For each analysis: Box's M , within-groups correlation matrix, within-groups covariance matrix, separate-groups covariance matrix, total covariance matrix. For each canonical discriminant function: eigenvalue, percentage of variance, canonical correlation, Wilks' lambda, chi-square. For each step, unstandardized function coefficients, Wilks' lambda for each canonical function.

The following discuss attempts to further explain some tables that can be obtained when the discriminate analysis procedures are carried out

Group statistics tables: The Group Statistics and Tests of Equality of Group Means tables provide information whether there are any significant differences between groups on each of the independent variables using group means and ANOVA results data.

Group centroids table: A further way of interpreting discriminant analysis results is to describe each group in terms of its profile, using the group means of the predictor variables. These group means are called centroids. Cases with scores near to a centroid are predicted as belonging to that group Venkatesh, (2018).

Log determinants and Box's M tables: DA the basic assumption is that the variance-covariance matrices are equivalent. Box's M tests the null hypothesis that the covariance matrices do not differ between groups formed by the dependent. For this assumption to hold, the log determinants should be equal.

Table of eigenvalues: This provides information on each of the discriminate functions (equations) produced. With only one function it provides an index of overall model fit which is interpreted as being the proportion of variance explained (R^2).

Wilks' lambda Table: Wilks' lambda indicates the significance of the discriminant function. This table indicates a proportion of total variability not explained, i.e. it is the converse of the squared canonical correlation.

The canonical discriminant function coefficient table: These unstandardized coefficients (b) are used to create the discriminant function (equation). The discriminant function coefficients b or standardized form beta both indicate the partial contribution of each variable to the discriminate function. They can be used to assess each independent variable's unique contribution to the discriminate function and therefore provide information on the relative importance of each variable. One can test the number of roots (function) that add *significantly* to the discrimination between groups. Only those found to be statistically significant are used for interpretation; non-

significant functions (roots) are to be ignored.

The standardized canonical discriminant function coefficients table: This table provides an index of the importance of each predictor like the standardized regression coefficients (beta's) does in multiple regression. The sign indicates the direction of the relationship.

The structure matrix table: It provides another way of indicating the relative importance of the predictors. As earlier stated, the factor structure coefficients are the correlations between the variables in the model and the discriminant functions. These Pearson coefficients are structure coefficients or discriminant loadings simply denote the simple correlations between the variables and the function(s). It can be used (interpreted), if one wants to assign substantive "meaningful" labels to the discriminant functions (akin to the interpretation of factors in factor analysis). They serve like factor loadings in factor analysis. By identifying the largest loadings for each discriminate function the researcher gains insight into how to name each function. Generally, just like factor loadings, 0.30 is seen as the cut-off between important and less important variables. It is mostly used in research because they are considered more accurate than the Standardized Canonical Discriminant Function Coefficients.

Classification table: Finally, there is the classification phase. The classification table, also called a confusion table, is simply a table in which the rows are the observed categories of the dependent and the columns are the predicted categories. When prediction is perfect all cases will lie on the diagonal. The percentage of cases on the diagonal is the percentage of correct classifications. (Venkatesh, 2018)

Types of Discriminate Analysis

According to Tabachnick & Fidell (2013), there are three main types of discriminate analysis. They are Standard (direct), sequential and stepwise discriminate analysis

Standard (Direct) Discriminate Analysis; In Standard (Direct) Discriminate Analysis, all predictors are entered in the equation at once and each predictor is assigned only the unique association it has with groups. Variance shared among predictors contribute to the total relationship but not so by any one predictor

Sequential (Hierarchical) Discriminate Analysis; Sequential (Hierarchical) Discriminate Analysis is used to evaluate contribution of predictors of group membership by predictors as they enter the equation in an order determined by the researcher. The researcher assess improvement in classification when a new predictor is added to a set of prior cases

Stepwise (Statistical) Discriminate Analysis; Stepwise (Statistical) Discriminate Analysis is used when the researcher has no reason for assigning some predictors higher priority than others. That is statistical criteria is used in determining the order of entry in preliminary research. That is the researcher wants a reduced set of predictors but has no set of preference among them. Probably the most common application of discriminant function analysis is to include many measures in the study, in order to determine the ones that discriminate between groups. Stepwise (statistical) discriminate analysis can be forward stepwise analysis and backward stepwise analysis.

Summary

Discriminant analysis is a statistical method that is used in research to predict group membership from a set of predictors. It is used to classify individuals into one of two or more alternative groups on the basis of a set of measurements. Discriminant analysis builds a predictive model for group membership. The model is composed of a discriminant function (or, for more than two groups, a set of discriminant functions). The discriminant function coefficients b or standardized form β both indicate the partial contribution of each variable to the discriminate function. They can be used to assess each independent variables unique contribution to the discriminate function. In order to derive substantive "meaningful" labels for the discriminant functions, one can also examine the factor structure matrix with the correlations between the variables and the discriminant functions as well as other statistics associated with discriminate analysis procedures.

References

- Garson, G. D. (2008). Discriminant function analysis. Retrieved on 15/6/2018 from <https://web.archive.org/web/20080312065328/http://www2.chass.ncsu.edu/garson/pA765/discrim.htm> .
- Hardle, W., Simar, L. (2007). *Applied Multivariate Statistical Analysis*. Springer Berlin Heidelberg. pp. 289–303.
- Hill, T. & Lewicki, P. (2007). *STATISTICS: Methods and Applications*. StatSoft, Tulsa, OK.
- Iweka, F.O (2017). *Educational Research Methods and statistical Techniques*. Chifas Nigeria, Omoku, Rivers State
- Lohnes, P.R.(1988). *Educational Research, methodology and measurement;An international handbook*, edited Keeves.J.P. Oxford, England: Pergamon press
- Ramayah T., Ahmad , N.H., Abdul-Halim,H., Zainal, M.S. & May-Chiun Lo, M.(2010) Discriminant analysis: An illustrated example. *African Journal of Business Management* Vol. 4(9), 1654-1667
- Poulsen, J & French, A. (2018).*Discriminant Function Analysis*. Retrieved on 14/6/2018 from

<http://www.statisticssolutions.com/discriminant-analysis>

Tabachnick B.G & Fidell L.S (2013) *Using Multivariate Statistics*, 6th edition. New Jersey:Person Edu. Inc

Tabachnick, B.G. and L.S. Fidell. (1996). *Using Multivariate Statistics*. Harper Collins College Publishers: New York.

Venkatesh p.2018, discriminate analysis; an overview.retrived on 15/6/2018 from https://www.ibm.com/support/knowledgecenter/en/SSLVMB_23.0.0/spss/base/idh_disc.html

Wikipedia (2018).Discriminate analysis; retrieved June 15 2018

Ying Li (2012) Discriminate Analysis: Two groups. Retrieved from. https://www.ibm.com/support/knowledgecenter/en/SSLVMB_23.0.0/spss/base/idh_disc.html