Application of Discriminant Analysis in the Classification of Food Security Status

A.A. ONOJA[1*], O.L. BABASOLA[2], V. Ojiambo[3]
[1]Pan African University, Institute for Basic Sciences, Technology and Innovation, Nairobi, donmaston09@gmail.com, P.O.Box 62000-00200, Nairobi-Kenya,
[2] Pan African University, Institute for Basic Sciences, Technology and Innovation, babasolaoluwatosin@yahoo.com, P.O.Box 62000-00200, Nairobi- Kenya
[3]Jomo Kenyatta University of Agriculture And Technology vojiambo@jkuat.ac.ke, P.O.Box 62000-00200, Nairobi-Kenya

Abstract
Vital information is usually lost during ordinal classification problems that incur misclassification error which affects predictions. In an attempt to minimize this error, this study investigates the effectiveness of adopting Linear Quadratic Discriminant Analysis method in the classification of ordinal dataset problem involving three group cases. In predictions of Food Security Status, there is a need to employ a powerful statistical tool that can correctly classify a household based on the Food Consumption Scores Profile indicator into "Poor", "Borderline" and "Acceptable". The approach was used to classify food security status of two counties in region of Kenya. The summary classification results showed that 89.9% of the original grouped cases were correctly classified while 89.1% of the cross-validation grouped cases were correctly classified. This approach can be employed by major International Organizations and Government of nations in their quest to minimize hunger and starvation all over the world.

Keywords: *Discriminant Analysis, Linear Quadratic Discriminant Analysis (LQDA), Ordinal classification, Food Security, Food Consumption Scores.*

1. Introduction

In order to understand certain patterns and identify the right group for easy predictions in categorical data, there is every need to employ certain Statistical techniques such as the Ordinal regression analysis, Principal Component analysis, etc. However some of these approaches have proven over time to be tedious in its classifications and validation approaches in dealing with ordinal categorical data like the classification and prediction of Food Security three proxy indicator variables via; Poor, Borderline and Acceptable. Therefore this paper employed the use of Discriminant Analysis to classify and predict Food Security Status of Households. (Brown 1984) opines that Discriminant Analysis (also referred to as Discriminant Function Analysis) is a powerful descriptive and classification approach that was developed by R.A Fisher in 1936 to mainly describe the characteristics that are peculiar to certain groups (known as descriptive discriminant analysis); and classify them into cases (that is individuals, subjects or participants) into pre-existing groups based on their similarities in the cases and other cases within the group (sometimes this is referred to as predictive discriminant analysis). Standard classification approaches for nominal classes can be applied to ordinal classification problems by ignoring the ordering information in the class attribute. Training samples of ordinal data set are labeled by a set of ranks, which exhibits an ordering among the different categories. However, huge information is lost during this analysis process that can potentially influence the predictive performance of a classifier. For instance, in studying the food security status of a household, there is every need to classify each household surveyed into three categories, prior to the Food Consumption Score (FCS). In doing such classification, ignoring the ordering class attribute, a household food security status may be misclassified and give rise to false

prediction of a household food security status. Therefore, this study seeks to utilize the Discriminant Analysis approach to address the issues of classification in ordinal data set. The main objective of this study is to determine the effectiveness of Linear Quadratic Discriminant Analysis (LQDA) approach in the classification of ordinal data set problem.

2. Literature Review

(Pinstrup-Andersen 2009; Barbosa and Nelson 2016) argues that the term "food security" has been used over time to mean different things. There are countless explanations affiliated to the concept of food security in literature over time have suggested that it is very useful to measure household and individual welfare, specifically if combined with estimates of household food acquisition and allocation behaviour. If nutritional security is the goal of interest, estimates of access to food should be combined with estimates of access to clean water and good sanitation. Anthrometric measures are likely to be more appropriate than food security estimates to target policies and programs to improved child nutrition. (Pinstrup-Andersen, 2009) opines that the concept of food security has been widely used at the household level as a measure of welfare and attempts have been made to make the concept operationally useful in the design, implementation, and evaluation of programs, projects and policies. A household is considered food secure if it has the ability to acquire the food needed by its members to be food secure. In order to determine the food security status of a household, World Food Programme (WFP) adopted the use of the seven-days household dietary diversity food frequency to classify households based on three ordinal indicator levels via; "Poor", "Borderline", and "Acceptable" this have helped decision makers like WHO, UNICEF, World Bank etc. to distributes relief materials in Internally Display Persons (IDP) camp, disaster etc. According (Rencher 2012) In discriminant analysis for several groups, the major focus is on finding a linear combinations of variables that best separate the k groups of multivariate observations. (Brown 1984) uses the discriminant analysis approach in healthcare to study familiar clinical situations. His results showed how DA might be relevant to health care decisions, especially classification decisions. (Fransens, Prins, and Gool 2003) combined the Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) approaches to introduce a novel classification approach, which combines the normal directions idea with Support Vector Machine classifiers. The two make a natural and powerful match, as SVs are located nearby, and fully describe the decision surfaces. (Fernandez 2002) introduces the Non-parametric discriminant methods based on non-parametric group-specific probability densities to evaluate the performance of a discriminant criterion which was attained by estimating probabilities of misclassification of new observations in the validation data.

3. Materials and Methods

Procedure for Data collections

Two research assistants were employed and trained with regards to technical know-how, ethical and behavioural approach during data collections in the field. The data was collected using the smart phones ODK toolbox. The use of questionnaire structured in modules were used to interview correspondents, the module of interest was titled "Food Security Situation" which consist of sub-modules on coping strategy and household dietary diversity in order to obtain the frequency of food dietary diversity consumption in the last seven-days and ascertain its status. Two stage sampling method was employed in the data collection approach. In the first phase, the Cluster sampling was used to isolate the

targeted communities, and then simple random sampling without replacement (SRSWR) was employed in the second stage to randomly selecting the households. A total of Seven-hundred and fifty two entries were recorded from one hundred and fifty households visited. Further cleansing of the data and analysis were done using the predictive analytics software (PASW) version 21.0.

Assumptions of Linear Discriminant Analysis

Discriminant analyses calculate the probability of group membership based on the series of independent predictor variables. The predictor variables will be measure on a scale level measurement while the dependent variables will be categorical. The following assumptions are necessary in order to explore the discriminant analysis approach:

- The dependent variable categories must be mutually exclusive
- The predictors are independent of one another and are normally distributed with absence of outliers
- There is no presence of multicollinearity among the predictors.
- The relationship between all pairs if groups are linear

Note: When dealing with a dependent variable with three levels, the use of multi-normal-logistic regression approach is adopted while with two levels the binary logistic regression is used.

Observations vectors in the samples:

$$\begin{matrix} y_{11} & y_{21} \\ y_{12} & y_{22} \\ \vdots & \vdots \\ y_{1n_1} & y_{2n_2} \end{matrix}$$ Are transformed to scars $$\begin{matrix} Z_{11} & Z_{21} \\ Z_{12} & Z_{22} \\ \vdots & \vdots \\ Z_{1n_1} & Z_{2n_2} \end{matrix}$$

Consider the means:

$$\bar{Z}_1 = \sum_{i=1}^{n_1} \frac{Z_{1i}}{n_1} = \boldsymbol{a}'\bar{y}_1$$

$$\bar{Z}_2 = \sum_{i=1}^{n_1} \frac{Z_{2i}}{n_2} = \boldsymbol{a}'\bar{y}_2$$

where:

$$\bar{y}_1 = \sum_{i=1}^{n_1} \frac{y_{1i}}{n_1}$$

and

$$\bar{y}_2 = \sum_{i=1}^{n_2} \frac{y_{2i}}{n_2}$$

so that the vector $\boldsymbol{a}$, is the maximized the standardize difference

$$\frac{(\overline{Z_1} - \overline{Z_2})}{S_Z} = \frac{[\boldsymbol{a}'(\bar{y}_1 - \bar{y}_2)]^2}{\boldsymbol{a}' S_{P1} \boldsymbol{a}} \quad (1)$$

This will be maximum if

$$a = S_{P1}^{-1}(\bar{y}_1 - \bar{y}_2)$$

$$a = S_{P1}^{-1}(\bar{y}_1 \bar{y}_2)$$

Note that $\boldsymbol{a}$ is not unique but its direction is unique. That us $\boldsymbol{a_1}, \boldsymbol{a_2}, \dots, \boldsymbol{a_p}$ are unique and $Z = \boldsymbol{a}'y$ project points y onto the line on which $\frac{(\bar{Z}_1 - \bar{Z}_2)^2}{S_Z^2}$ is maximum also for S_{P1}^{-1} to exist, then, $n_1 + n_2 - 2 > p$. For the K groups (samples), consider n_i observations in the i^{th}group, transforming each observation vector y_{ij} to obtain , $Z_{ij} = a'y_{ij}, \quad i = 1,2, \dots, k; j = 1,2, \dots., n_i$.

Consider the mean

$$\bar{Z}_i = a'\bar{y}_i$$

where,

$$\bar{y}_i = \sum_{i=1}^{n} \frac{y_{ij}}{n_i}$$

with the prior aim to maximize the separate vectors $\bar{Z}_1, \bar{Z}_2, \dots, \bar{Z}_k$. To express separation among $\overline{Z}_1, \bar{Z}_2, \dots, \bar{Z}_k$, extend the separate criterion so that the *k*-group case can be express as

$$\boldsymbol{a}'(\bar{y}_1 - \bar{y}_2) = (\bar{y}_1 - \bar{y}_2)'\boldsymbol{a},$$

then from equation (1),

$$\frac{(\bar{Z}_1 - \bar{Z}_2)^2}{S_Z^2} = \frac{[\boldsymbol{a}'(\bar{y}_1 - \bar{y}_2)]^2}{\boldsymbol{a}' S_{Pl} \boldsymbol{a}} = \frac{\boldsymbol{a}'(\bar{y}_1 - \bar{y}_2)(\bar{y}_1 - \bar{y}_2)'\boldsymbol{a}}{\boldsymbol{a}' S_{Pl} \boldsymbol{a}} \quad (2)$$

To extend equation (2) to *k*-groups, the concept of $\boldsymbol{H}$ matrix from multivariate analysis of variance (MANOVA) is adopted instead of

$$(\bar{y}_1 - \bar{y}_2)(\bar{y}_1 - \bar{y}_2)',$$

so that $\boldsymbol{E}$ is used to replace S_{Pl} to get

$$\Lambda = \frac{\boldsymbol{a'Ha}}{\boldsymbol{a'Ea}} \quad (3)$$

which can be further express as

$$\Lambda = \frac{SSH(Z)}{SSE(Z)}$$

where $SSH(Z)$ and $SSE(Z)$ are the between and within sum of squares for *Z*. Then, equation (3) can be express as $\boldsymbol{a'Ha}=\Lambda\ \boldsymbol{a'Ea} \Rightarrow \boldsymbol{a'(Ha - \Lambda Ea) = 0}$ (4). Where the value of Λ and $\boldsymbol{a}$ are derived from the solutions of equation (4) in a search for the values of $\boldsymbol{a}$ that leads to maximum Λ. The solution $\boldsymbol{a' = 0'}$ is not feasible since it gives $\Lambda = {}^{0}/_{0}$ in equation (3), other solutions are obtained from the form

$$\boldsymbol{(E^{-1}H - \Lambda I)a = 0} \quad (5)$$

The solutions of equation (5) are the Eigen values $\lambda_1, \lambda_2, \dots, \lambda_s$. Associated with the eigenvectors $\boldsymbol{a_1, a_2, \dots, a_s}$ of $\mathbf{E^{-1}H}$**.** Now consider the ranks $\lambda_1 > \lambda_2 > \cdots > \lambda_s$, the number of non-zero eigenvalues $\boldsymbol{s}$ is the rank of $\mathbf{H}$ which can be found as the smaller of $k-1$ or $\boldsymbol{p}$**.** The largest eigenvalue $\lambda_1$is the maximum value of $\Lambda = \frac{\boldsymbol{a'Ha}}{\boldsymbol{a'Ea}}$**,** and the coefficient vector that produces the maximum is the corresponding eigenvector $\boldsymbol{a}$**.** The means are:

$$\bar{Z}_1 = a_1'y$$

, where $\bar{Z}_1$ is the dimension that maximizes the means separated. Consider the eigenvectors $\boldsymbol{a_1, a_2, \dots, a_s}$ of $\mathbf{E^{-1}H}$ corresponding to $\lambda_1, \lambda_2, \dots, \lambda_s$, then the $\boldsymbol{s}$ discriminant functions $Z_1 = a_1'y$, $Z_2 = a_2'y$,…, $Z_s = a_s'y$

which shows the dimensions of difference among

$$\bar{y}_1, \bar{y}_2, \dots, \bar{y}_k$$

. This discriminant functions are uncorrelated, though not orthogonal ($a_i'a_j = 0,\ for\ i \neq j$) since $\mathbf{E^{-1}H}$ is not symmetric. The relative importance of each discriminant function Z_i can be determined by looking at its eigenvalue as proportional to the total:

$$\lambda_i \Big/ \sum_{j=1}^{s} \lambda_j$$

this is often used to show group dissimilarities. Usually the discriminant function of the smallest group is omitted.

Measure of Association

In order to determine the measure of association between the dependent variables $y_1, y_2, \dots, y_p$ and the independent grouping variables i relating to μ_i, $i = 1,2, \dots, k$, the Roy's Statistic θ is used to replace R^2-like measure of association. Since it is the ratio of between total sum of squares for the first discriminant

function

$$Z_1 = a_1' y,$$
$$\eta_\theta^2 = \theta = \frac{\lambda_1}{1+\lambda_1} = \frac{SSH(Z_1)}{SSE(Z_1) + SSH(Z_1)}$$

.

Canonical Correlation

(Brown 1984) opines that the canonical correlation is to measure variability by a function that is unrelated to the group differences. The variability may be related to within sum of squared group differences or to various errors that occurred during the data collection and entry. The canonical correlation shows the amount of group variability retained by each function while the canonical correlation squared indicates the proportion of variance in a function that is associated to group differences. The squared canonical correlation, can also be used to calculate for each discriminant function:

$$r_i^2 = \frac{\lambda_i}{1+\lambda_i}, \quad i = 1,2,\dots,s$$

. The average squared canonical correlation is used as measure of association. The larger the eigenvalues the more the correlations shows important functions.

Standardized Discriminant Functions

(Rencher 2012) used two group cases for the i^{th}observation vector y_{1i} *or* y_{2i} in group 1 or 2, to express the discriminant function in terms of standardized variables as:

$$Z_{1I} = \boldsymbol{a_1^*}\frac{y_{1i_1} - \bar{y}_{11}}{S_1} + \boldsymbol{a_2^*}\frac{y_{1i_2} - \bar{y}_{12}}{S_2} + \cdots + \boldsymbol{a_P^*}\frac{y_{1i_P} - \bar{y}_{1P}}{S_P} \qquad (8), i = 1,2,\dots,n_1$$

$$Z_{2I} = \boldsymbol{a_1^*}\frac{y_{2i_1} - \bar{y}_{21}}{S_1} + \boldsymbol{a_2^*}\frac{y_{2i_2} - \bar{y}_{22}}{S_2} + \cdots + \boldsymbol{a_P^*}\frac{y_{2i_P} - \bar{y}_{2P}}{S_P} \qquad (9), i = 1,2,\dots,n_2$$

Where $\bar{y}_1' = (\bar{y}_{11}, \bar{y}_{12}, \dots, \bar{y}_{1P})$ and $\bar{y}_2' = (\bar{y}_{21}, \bar{y}_{22}, \dots, \bar{y}_{2P})$ are the mean vectors for the two groups, and S_r is the within sample standard deviation of the r^{th} variable, obtained as the square root of the r^{th} diagonal element of S_{pl} so that $\boldsymbol{a_r^*} = S_r \boldsymbol{a_r}$, $r = 1,2,\dots,p$ in vector form, this becomes $\boldsymbol{a}^* = \left(diag S_{pl}\right)^{1/2}\boldsymbol{a}$. For several group cases, standardize the discriminant function in analogous form. The r^{th} dente the coefficient in the m^{th} discriminant function by a_{mr}, $m = 1,2,\dots,s$, $r = 1,2,\dots,p$ then the standardized form is:

$$\boldsymbol{a_{mr}^*} = S_r \boldsymbol{a_{mr}}$$

where S_r is the within-group standardized deviation obtained from the diagonal of

$$S_{p1} = \boldsymbol{E}/\boldsymbol{V_E}$$

$\boldsymbol{a_{mr}}$ have two subscripts due to the several discriminant functions.

Test of Significance

This is required in order to check for the assumption of multivariate normality.

The Wilk's Λ −test of significance

This is used to test for differences among mean vectors that are used. This is given by:

$$\Lambda_1 = \prod_{i=1}^{S} \frac{1}{1+\lambda_I},$$

which is distributed as $\Lambda_{p,k-1}$, where $N = \sum_{i=1} n_i$ for which unbalance design or $N = kn$ in balance case. Where Λ_1 is the small if one or more λ_i's are large, Wilk's tests for significance of the eigenvalue and thereby the discriminant functions. The *S* eigenvalues represent *S* dimensions of separation of the mean vectors $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_k$. Emphasize lie whether any of these dimensions are significant. More so, the exact test provided by the critical values for Λ using χ^2 −approaximation given by

$$V_E = N - k = \sum_{i=1} n_i - k$$

and

$$V_H = k - 1$$

$$V_1 = -\left[V_E - \frac{1}{2}(P - V_H + 1)\right] ln\Lambda_1 = -\left[N - 1 - \frac{1}{2}(P + k)\right] ln \prod_{i=1}^{S} \frac{1}{1+\lambda_I}$$

$$= -\left[N - 1 - \frac{1}{2}(P + k)\right] \sum_{i=1}^{S} ln(1+\lambda_i) \qquad (10)$$

, Which is approximately $\chi^2$with $p(k-1)$ degrees of freedom. If the test leads to rejection of H_0, then at least one of the Λ's is significantly different from zero, which shows that there is at least one dimension of separation of mean vectors. In general, for m^{th}step, the test statistic can be express as:

$$\Lambda_m = \prod_{i=m}^{S} \frac{1}{1+\lambda_I}$$

which is distributed as $\Lambda_{p-m+1,N-k-m+1}$, the statistic is given by:

$$V_m = -\left[N - 1 - \frac{1}{2}(P + k)\right] ln\Lambda_m = -\left[N - 1 - \frac{1}{2}(P + k)\right] \sum_{i=m}^{S} ln(1+\lambda_i) \qquad (11)$$

has an approximation as $\chi^2 -$ distribution with $(P + m + 1)(k - m)$ degrees of freedom. If

$$\lambda_i \Big/ \sum_j \lambda_j$$

is small, the associated discriminant function may not be of interest, even if it is significant.

Predictive Discriminant Analysis

Classification into several Groups: Equal Population Covariance Matrices Linear Classification Function. Assume that $\sum_1 = \sum_2 = \cdots = \sum_k$ the estimated common population covariance matrix given by the pooled sample covariance matrix as:

$$S_{pl} = \frac{1}{N-k}\sum_{i=1}^{k}(n_i - 1)S_i = \frac{\boldsymbol{E}}{\boldsymbol{N-k}} \qquad (12)$$

, where n_i and S_i are the sample size and covariance matrix of the i^{th} group, $\boldsymbol{E}$ is the error matrix for one-way *MANOVA* and $N = \sum_i n_i$, compare y to each $\bar{y}_i$, $i = 1,2,\dots,k$. The distance function is given by

$$D_i^2(y) = (y - \bar{y}_i)' {S_{pl}}^{-1}(y - \bar{y}_i) \qquad (13)$$

, and assign y to the group for which $D_i^2(y)$ is smallest. Expanding equation (13), the linear classification rule is derived as

$$D_i^2(y) = y'S_{pl}^{-1}y - y'S_{pl}^{-1}\bar{y}_i - \bar{y}_i'S_{pl}^{-1}y + \bar{y}_i'S_{pl}^{-1}\bar{y}_i = y'S_{pl}^{-1}y - 2\bar{y}_iS_{pl}^{-1}y + \bar{y}_i'S_{pl}^{-1}\bar{y}_i$$

. Neglect the term $y'S_{pl}^{-1}y$ on the *RHS* since it is a function of i. The second term is a linear function of y, and the third term does not have y. Therefore, omit $y'S_{pl}^{-1}y$ and obtain a linear classification function, denoted by $L_i(y)$. From normal distribution and prior probabilities, multiply through by $-1/2$, then assign y to the group, so that

$$L_i(y) = \bar{y}_i'S_{pl}^{-1}y - \frac{1}{2}\bar{y}_i'S_{pl}^{-1}\bar{y}_i \qquad ,i = 1,..,k, \qquad (14)$$

is a maximum. As a function of y, this can be written as:

$$L_i(y) = c_i'y + c_{io} = c_{i1}y_1 + c_{i2}y_2 + \cdots + c_{ip}y_p + c_{io}$$

where

$$c_i' = \bar{y}_i'S_{pl}^{-1}$$

and

$$c_{io} = -\frac{1}{2}\bar{y}_i'S_{pl}^{-1}\bar{y}_i$$

. Calculate c_i and c_{io} for each of the k-groups, evaluate $L_i(y)$, $i = 1,\dots,k$ and allocate y to the group for which $L_i(y)$ is the largest. Assign y to the group for which $p_i f(y|G_i)$ is maximum to minimize the probability of misclassification. Assume normality with equal covariance matrices and with probabilities of group membership $p_1, p_2, \dots, p_k$, then $f(y|G_i) = N_p(\mu_i, Z)$. Then the rule in equation (14) can be estimated in place of parameter

$$L_i(y) = lnp_i + \bar{y}_i{}'S_{pl}^{-1}y - \frac{1}{2}\bar{y}_i{}'S_{pl}^{-1}\bar{y}_i, \quad i = 1, \dots, k$$

and assign y to the group with maximum value of $L_i(y)$, where $p_1 = p_2 = \cdots = p_k$, $L_i(y)$ is the linear classification function. If

$$\sum_1 = \sum_2 = \cdots = \sum_k$$

does not hold, the classification rules can easily be altered to preserve optimality of classification. In place of equation (12), consider:

$$D_i^2(y) = (y - \bar{y}_i)'S_i{}^{-1}(y - \bar{y}_i) \qquad , i = 1, .., k, \quad (15)$$

Where S_i is the sample covariance matrix for the i^{th}group following same rule and approach the linear function of y is a quadratic function. Replace S_i with S_{pl} so that the optimal rule based on $p_i f(y|G_i)$ will now be:

$$Q_i(y) = lnp_i - \frac{1}{2}\ln|S_i| - \frac{1}{2}(y - \bar{y}_i)'S_i{}^{-1}(y - \bar{y}_i),$$

is maximum if $p_1 = p_2 = \cdots = p_k$ and $n_i > p$ for S_i^{-1} to exist.

Group Centroids

This refers to the mean discriminant scores of the members of a group on a given discriminant function. For classification and prediction purposes, the discriminant score of each group case is compare to each group centroids and the probability of group membership are obtained.

Estimating Misclassification Rates

In order to ascertain the power of classification approach so as to improve on accuracy of predictions of group membership correctly, there is every need to adopt the use of probability misclassification approach also known as the error rate. Whereas the complement of misclassification rate is refers to as correct classification rate. The proportion of misclassification resulting from resubstitution is known as apparent error rate. The results can be represented in a classification table also known as the confusion matrix. For two groups, the apparent error is given by:

$$Apparent\ error\ rate = \frac{n_{12} + n_{21}}{n_1 + n_2} = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}}$$

While the apparent correct classification rate

$$= \frac{n_{11} + n_{22}}{n_1 + n_2}$$

, thus, apparent error rate $= 1 -$ apparent correct classification rate

Cross-Validation

(Fernandez 2002) opines that this is a critical approach to verification of a discriminant analysis results especially if the researcher tends to classify other samples into the group of interest. In cross-validation, there are several methods but the most likely used methods include the Jack Knife procedure, the Hold-out method. The later approach involves splitting the sample randomly into two parts with two-thirds of the sample belonging to "developmental" sample and one-third of the sample allocated to a "cross-validation" sample. The accuracy of classification for the small sample indicates the hit rate the Researcher expects to achieve in fewer samples.

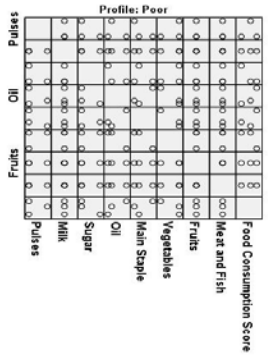
4. Results

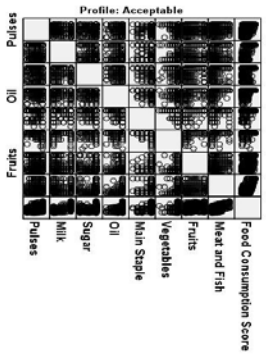
The data for this study was analyzed using the predictive analytics software popularly referred to as SPSS, the results were graphed and using tables to elicit some vital information.

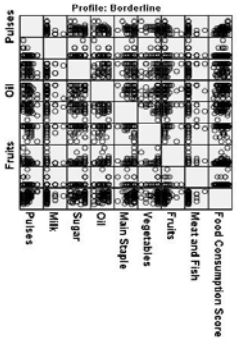
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Pulses	.232	752	.000	.859	752	.000
Milk	.310	752	.000	.715	752	.000
Sugar	.470	752	.000	.511	752	.000
Oil	.474	752	.000	.494	752	.000
Main Staple	.515	752	.000	.349	752	.000
Vegetables	.518	752	.000	.269	752	.000
Fruits	.233	752	.000	.772	752	.000
Meat and Fish	.250	752	.000	.752	752	.000
Food Consumption Score	.072	752	.000	.980	752	.000

a. Lilliefors Significance Correction







Correlations

		Pulses	Milk	Sugar	Oil	Main Staple	Vegetables	Fruits	Meat and Fish
Pulses	Pearson Correlation	1	.177**	.127**	.183**	.198**	.032	.060	.146**
	Sig. (2-tailed)		.000	.000	.000	.000	.377	.100	.000
	N	752	752	752	752	752	752	752	752
Milk	Pearson Correlation	.177**	1	.150**	.175**	.154**	.110**	.174**	.185**
	Sig. (2-tailed)	.000		.000	.000	.000	.002	.000	.000
	N	752	752	752	752	752	752	752	752
Sugar	Pearson Correlation	.127**	.150**	1	.330**	.115**	.213**	.141**	.145**
	Sig. (2-tailed)	.000	.000		.000	.002	.000	.000	.000
	N	752	752	752	752	752	752	752	752
Oil	Pearson Correlation	.183**	.175**	.330**	1	.293**	.377**	.195**	.238**
	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000	.000
	N	752	752	752	752	752	752	752	752
Main Staple	Pearson Correlation	.198**	.154**	.115**	.293**	1	.080*	.166**	.112**
	Sig. (2-tailed)	.000	.000	.002	.000		.028	.000	.002
	N	752	752	752	752	752	752	752	752
Vegetables	Pearson Correlation	.032	.110**	.213**	.377**	.080*	1	.101**	.094**
	Sig. (2-tailed)	.377	.002	.000	.000	.028		.005	.010
	N	752	752	752	752	752	752	752	752
Fruits	Pearson Correlation	.060	.174**	.141**	.195**	.166**	.101**	1	.344**
	Sig. (2-tailed)	.100	.000	.000	.000	.000	.005		.000
	N	752	752	752	752	752	752	752	752
Meat and Fish	Pearson Correlation	.146**	.185**	.145**	.238**	.112**	.094**	.344**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.002	.010	.000	
	N	752	752	752	752	752	752	752	752

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Group Statistics

Profile		Mean	Std. Deviation	Valid N (listwise) Unweighted	Weighted
Poor	Pulses	2.0000	3.46410	3	3.000
	Milk	.0000	.00000	3	3.000
	Sugar	.5000	.86603	3	3.000
	Oil	2.0000	1.32288	3	3.000
	Main Staple	9.3333	4.16333	3	3.000
	Vegetables	3.3333	1.15470	3	3.000
	Fruits	.0000	.00000	3	3.000
	Meat and Fish	.0000	.00000	3	3.000
Borderline	Pulses	6.0435	4.86716	69	69.000
	Milk	.3478	1.32650	69	69.000
	Sugar	2.3261	1.32794	69	69.000
	Oil	2.4710	1.10442	69	69.000
	Main Staple	11.6522	3.21680	69	69.000
	Vegetables	6.2899	1.68122	69	69.000
	Fruits	.4783	1.17083	69	69.000
	Meat and Fish	.3478	1.32650	69	69.000
Acceptable	Pulses	14.9647	6.37895	680	680.000
	Milk	15.9941	12.84231	680	680.000
	Sugar	3.0919	.99038	680	680.000
	Oil	3.2662	.62745	680	680.000
	Main Staple	13.7059	1.15205	680	680.000
	Vegetables	6.8471	.70355	680	680.000
	Fruits	2.1765	2.45334	680	680.000
	Meat and Fish	5.9765	7.40334	680	680.000
Total	Pulses	14.0944	6.79648	752	752.000
	Milk	14.4947	13.05891	752	752.000
	Sugar	3.0113	1.05958	752	752.000
	Oil	3.1882	.72764	752	752.000
	Main Staple	13.5000	1.61385	752	752.000
	Vegetables	6.7819	.88352	752	752.000
	Fruits	2.0120	2.41304	752	752.000
	Meat and Fish	5.4362	7.24398	752	752.000

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
Pulses	.844	69.442	2	749	.000
Milk	.875	53.345	2	749	.000
Sugar	.934	26.508	2	749	.000
Oil	.890	46.431	2	749	.000
Main Staple	.838	72.297	2	749	.000
Vegetables	.906	38.983	2	749	.000
Fruits	.956	17.278	2	749	.000
Meat and Fish	.947	20.800	2	749	.000

Pooled Within-Groups Matrices

		Pulses	Milk	Sugar	Oil	Main Staple	Vegetables	Fruits	Meat and Fish
Correlation	Pulses	1.000	.044	.035	.059	.048	-.074	-.026	.062
	Milk	.044	1.000	.075	.067	.018	.034	.109	.114
	Sugar	.035	.075	1.000	.272	.015	.150	.096	.097
	Oil	.059	.067	.272	1.000	.186	.327	.136	.176
	Main Staple	.048	.018	.015	.186	1.000	-.031	.093	.024
	Vegetables	-.074	.034	.150	.327	-.031	1.000	.055	.045
	Fruits	-.026	.109	.096	.136	.093	.055	1.000	.311
	Meat and Fish	.062	.114	.097	.176	.024	.045	.311	1.000

Log Determinants

Profile	Rank	Log Determinant
Poor	.[a]	.[b]
Borderline	8	7.265
Acceptable	8	12.786
Pooled within-groups	8	13.602

The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

a. Rank < 3

b. Too few cases to be non-singular

Test Results[a]

Box's M		1000.998
F	Approx.	26.630
	df1	36
	df2	47177.228
	Sig.	.000

Tests null hypothesis of equal population covariance matrices.

a. Some covariance matrices are singular and the usual procedure will not work. The non-singular groups will be tested against their own pooled within-groups covariance matrix. The log of its determinant is 13.624.

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.632[a]	93.0	93.0	.622
2	.048[a]	7.0	100.0	.213

a. First 2 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.585	399.733	16	.000
2	.954	34.778	7	.000

Standardized Canonical Discriminant Function Coefficients

	Function	
	1	2
Pulses	.504	-.130
Milk	.377	-.341
Sugar	.175	.313
Oil	.107	-.477
Main Staple	.496	.167
Vegetables	.317	.877
Fruits	.106	-.104
Meat and Fish	.122	-.118

Structure Matrix

	Function	
	1	2
Main Staple	.553*	.031
Pulses	.538*	-.224
Milk	.465*	-.348
Oil	.441*	-.139
Sugar	.327*	.266
Meat and Fish	.291*	-.208
Fruits	.267*	-.146
Vegetables	.350	.750*

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.

*. Largest absolute correlation between each variable and any discriminant function

Canonical Discriminant Function Coefficients

	Function	
	1	2
Pulses	.081	-.021
Milk	.031	-.028
Sugar	.170	.306
Oil	.156	-.694
Main Staple	.335	.113
Vegetables	.377	1.041
Fruits	.045	-.044
Meat and Fish	.017	-.017
(Constant)	-9.859	-6.421

Unstandardized coefficients

Functions at Group Centroids

	Function	
Profile	1	2
Poor	-4.915	-3.170
Borderline	-2.275	.282
Acceptable	.253	-.015

Unstandardized canonical discriminant functions evaluated at group means

Classification Processing Summary

Processed		752
Excluded	Missing or out-of-range group codes	0
	At least one missing discriminating variable	0
Used in Output		752

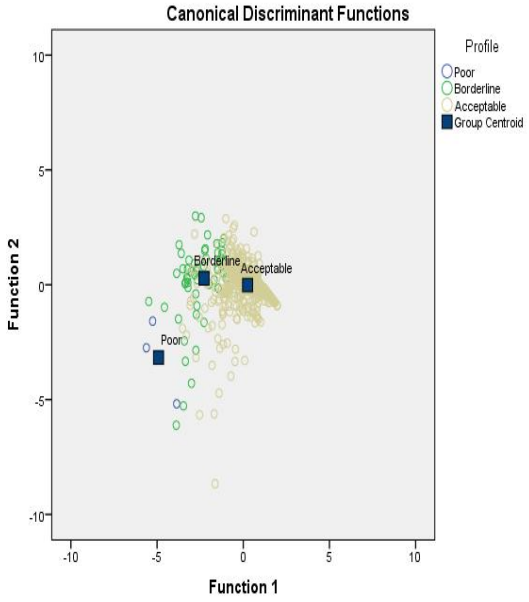
Prior Probabilities for Groups

		Cases Used in Analysis	
Profile	Prior	Unweighted	Weighted
Poor	.333	3	3.000
Borderline	.333	69	69.000
Acceptable	.333	680	680.000
Total	1.000	752	752.000

Classification Function Coefficients

	Profile		
	Poor	Borderline	Acceptable
Pulses	.046	.187	.396
Milk	-.015	-.030	.057
Sugar	-.282	1.222	1.562
Oil	.867	-1.117	-.518
Main Staple	4.323	5.599	6.413
Vegetables	4.851	9.440	10.084
Fruits	-.335	-.369	-.242
Meat and Fish	-.023	-.035	.013
(Constant)	-30.199	-63.911	-84.333

Fisher's linear discriminant functions



Classification Results[a,c]

		Profile	Predicted Group Membership			Total
			Poor	Borderline	Acceptable	
Original	Count	Poor	3	0	0	3
		Borderline	9	54	6	69
		Acceptable	8	54	618	680
	%	Poor	100.0	.0	.0	100.0
		Borderline	13.0	78.3	8.7	100.0
		Acceptable	1.2	7.9	90.9	100.0
Cross-validated[b]	Count	Poor	3	0	0	3
		Borderline	9	53	7	69
		Acceptable	9	57	614	680
	%	Poor	100.0	.0	.0	100.0
		Borderline	13.0	76.8	10.1	100.0
		Acceptable	1.3	8.4	90.3	100.0

a. 89.8% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 89.1% of cross-validated grouped cases correctly classified.

5. Discussion

The first table showed the test for normality. The P-Value for Shapiro-Wilk's test is needed to check if there are outliers and how well the data is normally distributed. The assumption however was violated since the p-values showed that they are all statistically significant. This however showed that the distribution was mostly skewed to the right-tailed and majority of the households surveyed were at the Borderline and Acceptable. Only few households were Poor. In order to check for linearity, the data file is split since it is organized based on group, then patterns are observed using the aid of the scatter plots from bottom left to the right upward. From the scatter plots it is clearly that the first plot has the dots scatter randomly, the second and third plots showed the dots mostly concentrated to the right. Though this is not in-line with the expectations of the Researcher, in general the assumption was slightly satisfied. Next the assumption of multicollinearity is considered by carrying out a bivariate correlation test. The Pearson correlation at 0.05 and 0.01 two-tailed level of significance tests were used to check for the weakest correlation 0.032 and the strongest correlation 0.585 this showed that all the correlations were

below 0.80 or 0.90 which means that the assumption of multicollinearity was satisfied. Having satisfied all the assumptions of prime importance, one can go ahead to carry out the linear quadratic discriminant analysis test. In the group statistics table among all the information listed, the mean for the three levels of the outcome variables and the predictor variables are of prime importance (consider from the least to the greatest). Under Poor Profile indicator mean levels, Milk, Fruits, Meat and Fish have the lowest 0.00, Main Staple had the highest 9.33, under Borderline indicator profile, , Milk, Fruits, Meat and Fish have the lowest 0.35, Main Staple had the highest 11.65, under Acceptable indicator profile, Fruits had the lowest 2.18 while Milk has the highest 15.99. In overall, Milk is the highest with 14.49 while Fruits was the lowest with 2.01. The next table to consider was the test of Equality group means table. The significant column showed that all the P-Values for the Wilk's Lambda are significant for the predictor variables. The Box's test of equality of covariance matrices tables looking mainly at the log-determinant column; the first row revealed that there are too few cases to be non-singular; the second row 7.265 is far close to 12.786 and 13.602 which are much closer to each other. From the test result table, it can be clearly seen that all the results are statistically significant which implies that they do not have equal covariance matrices. Note: here the p-Value is evaluated with $\alpha = 0.01$ not 0.05 which is commonly used. Consider the Canonical discriminant functions for the eigenvalues test, here considerations are given to the two listed with priority given to the function with greater eigenvalues which is 0.632 this means that the data fit the model. Consider the canonical correlation of the greatest eigenvalue which is 0.622, the square of this is 0.3869 this denote the effect size. Consider the Wilk's Lambda table shows that the both test functions are statistically significant, consider next the standardized canonical discriminant function coefficients table, compare this with the structure matrix table, ideally this is expected to be similar for instance Pulses is 0.504 in the latter table and 0.538 in the former table and so on. Though not all the values were consistent e.g. Milk is 0.377 in latter and 0.465 in the former which shows a lot of disparity, but the major goal here is to check if all the values in the structure matrix table are greater than 0.30. Next, consider the canonical discriminant function coefficients table, this is needed to build a discriminant function, the constant at the bottom is -9.859 then the unstandardized coefficients in function 1 are used. Consider the functions at Group Centroids, with main concentrations on function 1. The Profile column under Poor is -4.915, Borderline is -2.275 and Acceptable is 0.253 these are needed to make comparison between each group membership discriminant score and the probability of how each members were obtained. This is vital for classification and prediction purposes. Now consider the Classification Statistics, looking at the Classification coefficients table, the highest values here is Vegetables with Poor 4.85, Borderline 9.44, and Acceptable 10.08, second by Main Staple, Sugar and Pulses. Next, consider the Classification results table concentrating most on the percentage that the model accurately predicted the outcome. For Poor Profile indicator, 3 were counted and predicted 100%, Borderline counted 54 and predicted 78.3%, and Acceptable counted 618 and predicted 90.9%. Therefore, the highest level of classification is in the Poor Profile prediction which is 100%, the lowest is Borderline predictions which is 78.3%. The bottom summary classification result showed that 89.9% of the original grouped cases were correctly classified. While in the cross-validation analysis which is done mainly for the cases in the analysis, one can note that each case is classified by the functions derived from all cases other than that case itself. In this stance, 89.1% of cross-validated grouped cases were correctly classified. $Apparent\ error\ rate\ original\ classification = 1 - 0.899 = 0.101 = 10.1\%$.

6. Conclusion and Recommendations

In this study, it was shown that the Linear Quadratic Discriminant Analysis approach fit the model for classifications in large ordinal data set problem. From the results summary of classification it is worthy to note that most of the households in Kitui and Makueni regions of Kenya are generally food secure only few household have shown food insecurity. As such, this method can be used by international organizations like the World Food Programme (WFP), WHO, NEMA, UN, AU, etc. in their quest to conquer hunger and poverty in regions of Africa and the World. Linear Discriminant Analysis approach is a powerful classification approach which makes a good prediction yet it is worthy to note that as the data

set becomes bigger in size, the ability to adopt the use of Discriminant Analysis is meaningless, little wonder the error rate 10.1% and 10.9% these error margins are quite large and should not be taken slightly. Therefore, it will be recommended that a more powerful data mining tool be employ like the use of Support Vector Machine (SVM) classification approach that will incorporate the Kernel Method in the margin classification of support vectors along the hyperplanes, and since it perform better in high dimensional datasets.

Acknowledgement

We acknowledge the immense support of Dr. Ngesa Oscar, Prof. George Orwa, the entire Pan African University, Institute for Basic Sciences, Technology and Innovation (PAUISTI) and the Japan International Cooperation Agency (JICA), Kenya.

References

Barbosa, Rommel Melgaço, and Donald R. Nelson. 2016. "The Use of Support Vector Machine to Analyze Food Security in a Region of Brazil." *Applied Artificial Intelligence* 30 (4): 318–30. doi:10.1080/08839514.2016.1169048.

Brown, G W. 1984. "Discriminant Analysis." *American Journal of Diseases of Children (1911)* 138 (4): 395–400. doi:10.1016/B978-012691360-6/50009-4.

Fernandez, GCJ. 2002. "Discriminant Analysis, a Powerful Classification Technique in Data Mining." *Proceedings of the SAS Users International Conference*, 247–56. http://www2.sas.com/proceedings/sugi27/p247-27.pdf.

Fransens, Rik, Jan De Prins, and Luc Van Gool. 2003. "SVM-Based Nonparametric Discriminant Analysis, an Application to Face Detection." *Proceedings Ninth IEEE International Conference on Computer Vision*, no. Iccv: 1289–96 vol.2. doi:10.1109/ICCV.2003.1238639.

Pinstrup-Andersen, Per. 2009. "Food Security: Definition and Measurement." *Food Security* 1 (1): 5–7. doi:10.1007/s12571-008-0002-y.

Rencher, Alvin C. 2012. *Methods of Multivariate Analysis, Second Edition*. *IIE Transactions*. Vol. 37. doi:10.1080/07408170500232784.