International Journal of Multidisciplinary Research and Development
Online ISSN: 2349-4182, Print ISSN: 2349-5979; Impact Factor: RJIF 5.72
Received: 21-11-2018; Accepted: 24-12-2018
www.allsubjectjournal.com
Volume 6; Issue 1; January 2019; Page No. 170-176

Choose among Logistic, Probit regression and discriminant analysis in classification groups for processed food products in Bangladesh

Anwar Hossain[1*], Kalipada Sen[2], Nitai Chakraborty[3]
[1] Planning and Development Division, Bangladesh Council of Scientific and Industrial Research, Dr. Qudrat-I-Khuda Road, Dhanmondi, Dhaka-1205, Bangladesh.
[2, 3] Department of Statistics, Qazi Motahar Hossain Building (3rd floor),) University of Dhaka, Dhaka, Bangladesh

Abstract
Reliable with precedent studies, the logit and probit are similar under the different conditions. Hence, a selection between these two may not be significant (except in computational cost). However, the selection among logit, probit model, Linear Discriminant Analysis and Ordinary least squares (OLS) is still not uncomplicated. Therefore, the study has supposed to first carry out different preliminary data to decide the statistical properties of the forecaster variables. Possibly, part of the data might be analyzed by these techniques to decide which one is most suitable. Otherwise, the study could change the model to comply with the assumptions of a particular technique. From the results, the Logit and Probit Model perform the better results in terms of fulfill the assumptions. If in the case of assumptions are fulfill in Discriminant analysis yields better results than logit and probit model.

Keywords: logit and probit model, linear discriminant analysis, ordinary least squares (OLS)

Introduction

Popular methods used to analyze binary response data include the probit model, discriminant analysis, and logistic regression. Probit regression is based on the probability integral transformation. A major drawback of the probit model is that it lacks natural interpretation of regression parameters. Discriminant analysis is computationally simpler than the probit model. It assumes that predictor variables are normally distributed and that variables jointly assume a multivariate normal distribution. Because many variables in regression analysis are dichotomous or discrete, discriminant analysis assumptions are often violated. Furthermore, because discriminant analysis examines the distribution of X in terms of Y, it is dependent on Bayes theorem to extract the variable of primary interest. In contrast, the logistic regression model makes no assumption about the variable distribution. It is a direct probability model because it is stated in terms of $\Pr\{Y = 1|X\}$. Another advantage of the logit model is its ability to provide valid estimates, regardless of study design (Hailpern & Visintainer, 2003; Harrell, 2001; Newton *et al*., 2010) [6, 7, 16]. Ordinary least squares (OLS) regression, in its various forms, is the most common linear model analysis in the social sciences. If a dependent variable is a binary outcome, an analyst can choose among discriminant analysis and OLS, logistic or probit regression. OLS and logistic regression are the most common models used with binary outcomes (Pohlman & Leitner, 2003) [17]. To evaluate quality of industrially processed packed food products such as fortified high energy biscuit, chips, carbonated soft drink, corn flakes, Weetabix, juices and chocolate etc. in Bangladesh by measuring several parameters i.e. protein, fat, acidity, aflatoxin, iron, sugar, SPC, coliform etc. including pesticides for compliance of the BSTI/WFP/ICMSF standard in the framework of econometric tools. To use econometric model for assessing quality of industrially processed food products with especial reference to physiochemical and microbial analysis data.

Materials and Methods

The Binary Logistic Model

Logistic regression analysis is one of the most frequently used statistical procedures, and is especially common in medical research. The technique is becoming more popular in social science research (King & Ryan, 2002) [10].

Logistic regression estimates the probability of an outcome. Events are coded as binary variables with a value of 1 representing the occurrence of a target outcome, and a value of zero representing its absence. OLS can also model binary variables using linear probability models (Menard, 1995) [15]. OLS may give predicted values beyond the range (0, 1), but the analysis may still be useful for classification and hypothesis testing. The normal distribution and homogeneous error variance assumptions of OLS will likely be violated with a binary dependent variable, especially when the probability of the dependent event varies widely. Both models allow continuous, ordinal and/or categorical independent variables (Pohlman & Leitner, 2003) [17].

The dependent variable in logistic regression is usually dichotomous, that is, the dependent variable can take the value 1 with a probability of success θ, or the value 0 with probability of failure $1-\theta$. This type of variable is called a Bernoulli (or binary) variable. Although not as common and not discussed in this treatment, applications of logistic regression have also been extended to cases where the dependent variable is of more than two cases, known as multinomial or polytomous (Tabachnick & Fidell, 1996) [18].

As mentioned previously, the independent or predictor variables in logistic regression can take any form. That is, logistic regression makes no assumption about the distribution of the independent variables. They do not have to be normally distributed, linearly related or of equal variance within each group. The relationship between the predictor and response variables is not a linear function in logistic regression; instead, the logistic regression function is used, which is the logit transformation of θ:

$$\theta = \frac{e^{(\alpha+\beta_1 x_1+\beta_2 x_2+\cdots\beta_k x_k)}}{1+e^{(\alpha+\beta_1 x_1+\beta_2 x_2+\cdots\beta_k x_k)}}$$

Where α = the constant of the equation and, β = the coefficient of the predictor variables.
An alternative form of the logistic regression equation is:

$$\boldsymbol{logit[\theta(x)] = log\left[\frac{\theta(x)}{1-\theta(x)}\right] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_k x_k}$$

The goal of logistic regression is to correctly predict the category of outcome for individual cases using the most parsimonious model. To accomplish this goal, a model is created that includes all predictor variables that are useful in predicting the response variable. Several different options are available during model creation. Variables can be entered into the model in the order specified by the researcher or logistic regression can test the fit of the model after each coefficient is added or deleted, called stepwise regression (Interval & Ratio, 2016) [9].
Stepwise regression is used in the exploratory phase of research but it is not recommended for theory testing (Menard, 1995) [15]. Theory testing is the testing of a priori theories or hypotheses of the relationships between variables. Exploratory testing makes no a-priori assumptions regarding the relationships between the variables, thus the goal is to discover relationships.
Backward stepwise regression appears to be the preferred method of exploratory analyses, where the analysis begins with a full or saturated model and variables are eliminated from the model in an iterative process. The fit of the model is tested after the elimination of each variable to ensure that the model still adequately fits the data. When no more variables can be eliminated from the model, the analysis has been completed.
There are two main uses of logistic regression. The first is the prediction of group membership. Since logistic regression calculates the probability or success over the probability of failure, the results of the analysis are in the form of an odds ratio. For example, logistic regression is often used in epidemiological studies where the result of the analysis is the probability of developing cancer after controlling for other associated risks. Logistic regression also provides knowledge of the relationships and strengths among the variables (e.g., smoking 10 packs a day puts you at a higher risk for developing cancer than working in an asbestos mine).
The process by which coefficients are tested for significance for inclusion or elimination from the model involves several different techniques. Each of these will be discussed below.

Wald Test: A Wald test is used to test the statistical significance of each coefficient (β) in the model. A Wald test calculates a z statistic, which is:

$$z = \frac{\hat{B}}{SE}$$

This z value is then squared, yielding a Wald statistic with a chi-square distribution. However, several authors have identified problems with the use of the Wald statistic. (Menard, 1995) [15] Warns that for large coefficients, standard error is inflated, lowering the Wald statistic (chi-square) value. (Agresti, 1996) [1] States that the likelihood-ratio test is more reliable for small sample sizes than the Wald test.

Likelihood-Ratio Test: The likelihood-ratio test uses the ratio of the maximized value of the likelihood function for the full model (L_1) over the maximized value of the likelihood function for the simpler model (L_0). The likelihood-ratio test statistic equals:

$$-2\log\left(\frac{L_0}{L_1}\right) = -2[\log(L_0) - \log(L_1)] = -2(L_0 - L_1)$$

This log transformation of the likelihood functions yields a chi-squared statistic. This is the recommended test statistic to use when building a model through backward stepwise elimination (Agresti, 1996; Interval & Ratio, 2016) [3, 9].

The Binary Probit model

While Linear Probability Model (LPM) has a number of shortcomings that make it unsuitable; it is can generate probability values that lie below zero or above one, which would be unrealistic. LPM also leads to questionable values of R^2 as a measure of goodness of fit (Gujarati, 2003) [5]. This study assumes a normal cumulative distribution function and hence the choice of probit.
Binary response models are used when the number of alternatives that can be chosen is more than one. They are developed to describe the probability of each of the possible outcomes as a function of personal or alternative specific characteristics (Verbeek, 2008) [20]. Binary response models are applied where a binary or logical ordinal of the alternatives. In this case it is assumed that there exists an underlying latent variable that drives the choice between the alternatives (Verbeek, 2008) [20]. The results in this case will be sensitive to the way in which the alternatives are numbered. The modeling methodology used to establish the determinants of the quality of food products status is the binary probit model.
The binary probit is suitable for modeling with a categorical dependent variable (in this study the acceptable range of physiochemical analysis of food products status). Multivariate modeling is an especially useful and informative approach for understanding the accepted or unaccepted food products decision on their physiochemical analysis status. This is because multiple factors contribute to their decision on whether to be fully accepted or unaccepted. Binary probit is especially appropriate in this study because like Ordinary Least Square (OLS) it identifies the statistical significant relationships between the explanatory variables and the dependent variable. Unlike the OLS regression, binary probit discerns unequal differences between binary categories in the dependent variable (Greene, 2003; McKelvey & Zavoina, 1975) [4, 14].

Mathematical representation of the binary probit model

In this study, the dependent variable of the physiochemical analysis of food products status was placed in two categories. The food products are classified as fully accepted or unaccepted. A binary probit model is used to determine the qualitative acceptability of food products status. Based on

this, the model is estimated as follows:
Accepted or unaccepted food products status = f (physiochemical and microbial analyzed parameters of food products)
The food products status is modeled using the binary probit model with the model outcomes:
$S_i = 0$ (fully accepted) and
$S_i = 1$ (unaccepted).
The decision on food products status is unobserved and is denoted by the latent variable s_i^*. The latent equation below models how s_i^* varies with physiochemical analyzed parameter characteristics and is represented as:

$$s_i^* = X_i'\alpha + \varepsilon_i$$

Where, the latent variable s_i^*measures the difference in utility derived by individual i from either being fully-certified accepted or unaccepted. (*i*= 1, 2, 3…………….n), n represents the total number of food products samples. Each individual i belongs to one of the two groups.
X_i is a vector of exogenous variables,
α is a conformable parameter vector and the error term ε_i is independent and identically distributed as standard normal, that is $\varepsilon_i \sim NID(0, 1)$.
The observed variable (S_i) relates to the latent variable (s_i^*) such that

$$S_i = \begin{cases} 1 \text{ if } s_i^* \le 0 \\ 0 \text{ if } s_i^* > \gamma \end{cases}$$

Taking the value of 0 if the individual was fully-certified acceptable and 0 if the individual was unacceptable. The implied probabilities are obtained as:

$$\Pr\{S_i = 1 | X_i\} = \Pr\{s_i^* \le 0 | X_i\} = \Phi(-X_i'\alpha),$$

$$\Pr\{S_i = 3 | X_i\} = \Pr\{s_i^* > \gamma | X_i\} = 1 - \Phi(\gamma - X_i'\alpha)$$

And

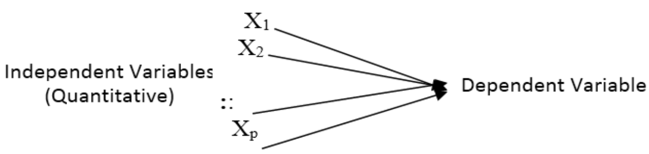
$$\Pr\{S_i = 2 | X_i\} = \Phi(\gamma - X_i'\alpha) - \Phi(-X_i'\alpha)$$

Where γ is the unknown parameter that is estimated jointly with α. Estimation is based upon the maximum likelihood where the above probabilities enter the likelihood function. The interpretation of α coefficient is in terms of the underlying latent variable model in equation.
The probability of the food products being fully-certified acceptable can be written as

$$\Pr(S_i = 1) = \Phi(X_i'\alpha_1)$$

Where Φ (.) is the cumulative distribution function (cdf) of the standard normal (Kisaka-Lwayo *et al.*, 2014; Verbeek, 2008) [11, 20].

Discriminant analysis model
Discriminant analysis is a statistical technique designed to investigate the differences between two or more groups of cases with respect to several underlying variables. This technique is more appropriate than commonly used measures (correlation, regression, etc) because the variables being predicted are categorical. Moreover, this approach results in a unit of analysis, predicted category membership that is more useful in evaluating instructional interventions. Its goal is to classify cases into one or several mutually exclusive groups based on their values for a linear combination of predictor variables (Manly, 1986; Thomas, Marr, Thomas, Hume, & Walker, 1996) [13, 19]. In this study, the foods are classified into two main groups: those that have never acceptable range of food products and those that have considered acceptable range of food products depending on their physiochemical analysis of food products status.
Since there are two groups, the number of unique functions that can be extracted is equal to (g-1), where g is the number of groups, or, p, the number of discriminant variables, whichever is less. In this study a Linear Discriminant Functions (LDF_s) are computed. The analysis assumes that the discriminant function scores (D_{km}) are normally distributed for each group and that the groups have equal variance covariance matrices for the discriminating variables. In practice these conditions are seldom applied strictly as the technique is very robust to departures from these assumptions (Klecka, 1980) [12].
In multiple linear regressions, the objective is to model one quantitative variable (called the dependent variable) as a linear combination of others variables (called the independent variables). The purpose of discriminant analysis is to obtain a model to predict a single qualitative variable from one or more independent variable (s). In most cases the dependent variable consists of two groups or classifications, like, high versus normal blood pressure, loan defaulting versus non-defaulting, use versus nonuse of internet banking etc. The choice between three candidates, A, B or C in an election is an example where the dependent variable consists of more than two groups (Data, Using, & Lesson, n.d.).
Discriminant analysis derives an equation as linear combination of the independent variables that will discriminate best between the groups in the dependent variable. This linear combination is known as the discriminant function. The weights assigned to each independent variable are corrected for the interrelationships among all the variables. The weights are referred to as discriminant coefficients.
The model:



The discriminant equation:

$$F = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$

where, F is a latent variable formed by the linear combination of the dependent variable, $X_1, X_2, \ldots X_p$ are the p independent variables, ε is the error term and $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ are the discriminant coefficients. The objective discriminant analysis is to test if the classifications of groups in a variable Y depend on at least one of the X_i's (Ideas & Notes, n.d.).

Assumptions: The assumptions of discriminant analysis are the same as those for MANOVA. The analysis is quite sensitive to outliers and the size of the smallest group must be larger than the number of predictor variables.

- The variables X_1, X_2,…, X_p are independent of each other.
- Groups are mutually exclusive and the group sizes are not grossly different.
- The number of independent variables is not more than two less than the sample size.
- The variance-covariance structures of the independent variables are similar within each group of the dependent variable.
- Errors (residuals) are randomly distributed.
- For purposes of significance testing, the independent variables follow a multivariate normal distribution (Data *et al*., n.d.).

There are several purposes for Multiple Discriminant Analysis (MDA):

- To investigate differences among groups.
- To determine the most parsimonious way to distinguish among groups.
- To discard variables which are little related to group distinction.
- To classify cases into groups.
- To test theory by observing whether cases are classified as predicted (Data *et al*., n.d.; Garson, 2009).

Data

Food products analysis data has been collected from Institute of Food Science and Technology (IFST), BCSIR with permission of the authority. The variable name of study are microbial count (SPC, Coliform, Mold, Yeast, E.Coli, Salmonella etc.), physical properties (Broken, Damaged, Moisture, Ash, Milling degree, Paddy etc.), chemical properties (Protein, Fat, Fibre, Carbohydrate, Energy, Acidity, Sugar etc.) and toxicity (Aflatoxin) and relevant data have been collected from the adhoc analysis report of different food products which are previously analyzed by respective scientist of IFST (Institute of Food Science and Technology (IFST), BCSIR, 2010) [8]. Data collection methods were non-participant observation of organization included in the study. Archival research included hard-copy issues of reports of analytical documents. We sought to collect each data over the five year period from 2007 to 2012 on a Single Stage Cluster Sampling basis.

Software

Statistical Analysis System STATA (Stata Corporation, College Station, TX, version 12) programs were used to analyze the data.

Results and Discussion

Choose among Logistic & Probit Regression and Discriminant Analysis in classification groups for fortified high energy biscuits.

Table 1: Summary of statistics of Logit, Probit model and Discriminant analysis.

		Logistic Regression		Probit Regression		Discriminant analysis	
Dependent Variable	**Independent Variable**	**p-value**	**GOF**	**p-value**	**GOF**	**p-value**	**GOF**
Protein (%)	Moisture (%)	0.101	0.112	0.103	0.112	0.102	0.751
Fat (%)		0.133	0.522	0.131	0.522	0.132	0.687
Sugar (%)		0.057	0.164	0.063	0.083	0.054	0.618
Total Carbohydrate (%)		0.932	0.115	0.933	0.115	0.932	0.000
Iron (mg/100g)		0.189	0.273	0.186	0.273	0.187	0.670
Vitamin A (mcg/100g)		0.123	0.460	0.122	0.460	0.121	0.027
Enterobacter sakazakii (cfu/g)		0.992	0.999	0.991	0.999	0.992	0.596
Yeast and moulds (cfu/g)		0.152	0.988	0.154	0.996	0.125	0.021

Note: GOF= Goodness-of-fit statistics.

From the above demonstrations of three different technique, Logit & Probit model and Discriminant analysis, all of them do not provide equal with the level of accepted range as prescribed by WFP, Dhaka predicted probability of the same variable which is given. The level of significance of Goodness-of-fit statistics are >0.05 under Logit and Probit, but under Discriminant analysis are >0.05 except Total Carbohydrate (%), Vitamin A (mcg/100g) and Yeast & Moulds count (cfu/g) as prescribed by WFP, Dhaka. Obviously, from these results, the Logit and Probit Model perform is preferable results in terms of fulfilling the assumptions. If in the case of normality assumptions are fulfill, Discriminant analysis also yields better results.

Table 2: Choose of statistical technique among Logit, Probit model and Discriminant analysis.

		Logistic Regression		Probit Regression		Discriminant analysis	
Responding variable	**Independent Variable**	**p-value**	**GOF**	**p-value**	**GOF**	**p-value**	**GOF**
Yeast and Mold Count (cfu/ml)	pH	0.395	0.70	0.360	0.71	0.221	-
	Total Soluble Solid (%)	0.647		0.628		0.391	
	Acidity (%)	0.455		0.443		0.225	
	Gas Pressure (lb/in^2)	0.349		0.328		0.162	
Standard Plate Count (cfu/ml)	pH	0.498	0.52	0.449	0.58	0.340	-
	Total Soluble Solid (%)	0.701		0.751		0.515	
	Acidity (%)	0.224		0.209		0.384	
	Gas Pressure (lb/in^2)	0.217		0.196		0.811	
Total Sugar (%)	pH	0.009	0.90	0.002	0.98	0.000	0.000
	Acidity (%)	0.005		0.002		0.003	
	Gas Pressure (lb/in^2)	0.329		0.222		0.151	

Note: GOF= Goodness-of-fit statistics.

From the above demonstrations of three different technique, Logit & Probit model and Discriminant analysis, all of them provide do not equal predicted probability of the same variable as prescribed by BSTI, Dhaka. The level of significance of Goodness-of-fit statistics are >0.05 under Logit and Probit, but under Discriminant analysis are <0.05 according to accepted range of Total Sugar (%). Obviously, from these results, the Logit and Probit Model perform the best results. If in the case of rejected predicted probability (p-value) Discriminant analysis yields same results as logit and probit model.

Table 3: Choose of statistical technique among Logit, Probit model and Discriminant analysis.

		Logistic Regression		Probit Regression		Discriminant analysis	
Responding variable	**Independent Variable**	**p-value**	**GOF**	**p-value**	**GOF**	**p-value**	**GOF**
Protein (%)	Moisture (%)	0.198	0.273	0.190	0.274	0.195	0.255
Total Ash (on dry basis) (%)		0.034	0.734	0.022	0.763	0.001	0.388
Tritratable Acidity (as lactic acid) (%)		0.709	0.318	0.687	0.325	0.725	0.089
Solubility (%)		0.547	0.281	0.554	0.282	0.565	0.407

Note: GOF= Goodness-of-fit statistics.

From the above demonstrations of three different technique, Logit & Probit model and Discriminant analysis, all of them provide are not exact equal predicted probability of the same variable which is given with the level of accepted range as prescribed by standard institution. The level of significance of Goodness-of-fit statistics are >0.05 under Logit & Probit and Discriminant analysis.

Table 4: Choose of statistical technique among Logit, Probit model and Discriminant analysis.

		Logistic Regression		Probit Regression		Discriminant analysis	
Responding variable	**Independent Variable**	**p-value**	**GOF**	**p-value**	**GOF**	**p-value**	**GOF**
Broken (%)	Moisture (%)	0.537	0.636	0.531	0.665	0.559	-
Damaged/discoloured (%)		0.977	0.319	0.977	0.319	0.979	0.451
Standard Plate Count (cfu/g)		0.180	0.389	0.170	0.390	0.132	0.056
Total Coli Form (MPN/g)		0.131	0.152	0.094	0.214	0.016	0.011

Note: GOF= Goodness-of-fit statistics.

From the above demonstrations of three different technique, Logit & Probit model and Discriminant analysis, all of them provide are not exact equal predicted probability of the same variable which is given with the level of accepted range as prescribed by standard institution. The level of significance of Goodness-of-fit statistics are >0.05 under Logit & Probit and <0.05 of Total Coli Form (MPN/g) under Discriminant analysis. Obviously, from these results, Logit &Probit model performs the best results in terms of fulfill the assumptions. If in the case of assumptions are fulfill in Discriminant analysis yields better results than logit and probit model.

Table 5: Choose among Logit, Probit model and Discriminant analysis.

		Logistic Regression		Probit Regression		Discriminant analysis	
Responding variable	Independent Variable	p-value	GOF	p-value	GOF	p-value	GOF
Acceptability of protein (%)	Moisture (%)	0.434	0.795	0.416	0.828	0.015	-
	Sugar (as sucrose) (%)	0.294		0.275		0.729	
	Total Carbohydrate (%)	0.579		0.566		0.017	
Acceptability of Fat (%)	Moisture (%)	0.421	0.429	0.419	0.433	0.005	0.061
	Sugar (as sucrose) (%)	0.293		0.289		0.687	
	Total Carbohydrate (%)	0.280		0.264		0.003	
Acceptability of Iron (mg/100g)	Moisture (%)	0.482	0.326	0.518	0.361	0.937	0.229
	Sugar (as sucrose) (%)	0.071		0.058		0.243	
	Total Carbohydrate (%)	0.190		0.160		0.054	

Note: GOF= Goodness-of-fit statistics.

From the above demonstrations of three different technique, Logit & Probit model and Discriminant analysis, all of them provide are not equal predicted probability of the same variable which is given with the level of accepted range as prescribed by WFP, Dhaka. The level of significance of Goodness-of-fit statistics are >0.05 under Logit & Probit and Discriminant analysis. Obviously, from these results, Discriminant analysis perform the better results in terms of the fulfill the assumptions except the dependent variable acceptability of protein (%) as there exist fewer than two nonsingular group covariance matrices. We know that if in the case of assumptions are fulfill in Discriminant analysis yields better results than logit and probit model.

Table 6: Choose among Logit, Probit model and Discriminant analysis.

		Logistic Regression		Probit Regression		Discriminant analysis	
Response variable	**Independent Variable**	**p-value**	**GOF**	**p-value**	**GOF**	**p-value**	**GOF**
Purity (%)	Moisture (%)	0.524	0.336	0.534	0.339	0.539	0.221
Heat damage (%)		0.303	0.437	0.304	0.446	0.309	0.730
Other damage (%)		0.215	0.624	0.203	0.617	0.206	0.732
Foreign matter (%)		0.751	0.422	0.761	0.421	0.761	0.386
Insect damage (%)		0.513	0.316	0.503	0.315	0.528	0.164
Broken (%)		0.317	0.655	0.296	0.664	0.322	0.492

Note: GOF= Goodness-of-fit statistics.

From the above demonstrations of three different technique, Logit & Probit model and Discriminant analysis, all of them provide almost equal predicted probability of the same variable which is given with the level of accepted range as prescribed by WFP, Dhaka. The level of significance of Goodness-of-fit statistics are >0.05 under Logit, Probit and Discriminant analysis.

Table 7: Choose among Logit, Probit model and Discriminant analysis.

		Logistic Regression		Probit Regression		Discriminant analysis	
Response variable	**Independent Variable**	**p-value**	**GOF**	**p-value**	**GOF**	**p-value**	**GOF**
Colour of the solution, in ICUMSA units	Moisture (%)	0.921	0.995	0.908	0.995	0.934	0.000
	Sucrose (%)	0.501		0.501		0.762	
Sulphated Ash (%)	Moisture (%)	0.038	1.000	0.058	1.000	0.010	-
	Sucrose (%)	0.370		0.459		0.925	

Note: GOF= Goodness-of-fit statistics.

From the above demonstrations of three different technique, Logit & Probit model and Discriminant analysis, all of them provide are not equal predicted probability of the same variable which is given with the level of accepted range as prescribed by WFP, Dhaka. The level of significance of Goodness-of-fit statistics are >0.05 under Logit and Probit, respectively but all variables under Discriminant analysis are <0.05. Obviously, from these results, the Logit and Probit Model perform the better results in terms of fulfill the assumptions. If in the case of assumptions are fulfill in Discriminant analysis yields better results than logit and probit model.

Acknowledgments
I would like to thank Authority of BCSIR for gives an opportunity to use of analytical service data and Ministry of Science and Technology to allocate research fund.

Conclusion

From the demonstrations of three different technique, Logit & Probit model and Discriminant function analysis, all of them provide are not exact equal predicted probability of the same variable which is given with the level of accepted range as prescribed by BSTI/WFP/ICMSF. From the results, Logit & Probit and Discriminant analysis of the level of significance of Goodness-of-fit statistics are >0.05 then, we conclude that Logit & Probit model performs the best results in terms of fulfill the assumptions. If in the case of assumptions fulfill in Discriminant analysis yields better results than logit and probit model.

Reverences

1. Agresti A. An Introduction to Categorical Data Analysis. New York: John Wiley & Sons, Inc, 1996.
2. Data M, Using A, Lesson S. (n.d.). Lesson 2: 27-42.
3. Garson GD. Discriminant Function Analysis, 2009, 2.
4. Greene WH. Econometric analysis. Pearson Education India, 2003.
5. Gujarati DN. Basic Econometrics. 4th. New York: McGraw-Hill, 2003.
6. Hailpern SM, Visintainer PF. Odds ratios and logistic regression: further examples of their use and interpretation. Interpretation. 2003; 318(134):0-356.
7. Harrell FE. Regression Modeling Strategies Springer Verlag. New York, NY, 2001.
8. Ideas M, Notes T. (n.d.). Lesson 2. In History Institute of Food Science and Technology (IFST), BCSIR, D. Approval for adhoc data use. Institute of Food Science and Technology (IFST), BCSIR, Dhaka, 2010, 14-23.
9. Interval C, Ratio O. Logistic Regression, 2016, 6-8.
10. King EN, Ryan TP. A preliminary investigation of maximum likelihood logistic regression versus exact logistic regression. The American Statistician. 2002; 56(3):163-170.
11. Kisaka-Lwayo M, Obi A, Obi MKL, Kisaka-Lwayo M, Obi A, Obi MKL. Analysis of production and consumption of organic products in South Africa. Chapter on Creative Commons Attribution, 2014, 1-26. https://doi.org/10.5772/58356
12. Klecka WR. Discriminant analysis. Sage, 1980.
13. Manly B. Multivariate statistical methods, a primer (1st ed.). London, UK: Chapman and Hall, 1986.
14. McKelvey RD, Zavoina W. A statistical model for the analysis of ordinal level dependent variables. Journal of Mathematical Sociology. 1975; 4(1):103-120.
15. Menard S. Applied logistic regression analysis: Sage university series on quantitative applications in the social sciences. Thousand Oaks, CA: Sage, 1995.
16. Newton HJ, Baum CF, Beck N, Cameron C, Epstein D, Hardin J, Kohler U. The Stata Journal. Stata Journal. 2010; 10(3):288-308. https://doi.org/The Stata Journal

17. Pohlman JT, Leitner DW. A comparison of ordinary least squares and logistic regression, 2003.
18. Tabachnick BG, Fidell LS. Using Multivariate Statistics, New York: HarperCollins College Publishers, 1996.
19. Thomas EW, Marr MJ, Thomas A, Hume RM, Walker N. Using discriminant analysis to identify students at risk. In Frontiers in Education Conference, 1996. FIE'96. 26th Annual Conference, Proceedings of. IEEE, 1996; 1:185-188.
20. Verbeek M. A guide to modern econometrics. John Wiley & Sons, 2008.