



## Comparison of discriminant analysis and logistic regression for the classification of borrowers of microfinance institutions

Chi Collins Penn

Department of Accountancy and Finance, Faculty of Economics and Applied Management, University of Douala, Cameroon  
Laboratory of Economics and Applied Management, University of Douala, Cameroon

### Abstract

This paper uses logistic regression and discriminant analysis to predict default in the loan portfolio of Microfinance institutions. Out of the 723 borrowers the logistic regression correctly classified 92.7% of good borrowers and 73.1% of bad borrowers. As for the discriminant analysis model it correctly classified 94.2% of the good borrowers and 84.5% of bad borrowers. The overall predictive accuracy of both methods was good enough with the logistic regression having the highest value of 81.6%. The results suggest that both methods can be used in selecting borrowers in the loan portfolio of microfinance institutions. Nevertheless given that the discriminant analysis does not meet up with its underlying assumptions, logistic regression is recommended.

**Keywords:** predict, default, loan, portfolio, classification, microcredit

### Introduction

Discriminant analysis and logistic regression can be used to predict the probability of a specified outcome such as default in a loan portfolio using all or a set of available variables. The two methods can also be used to examine the effects of each variable on loan default (or the change in the risk of loan default due to any of the variables) while controlling the effects of, and interrelationships with, other variables. Logistic regression is becoming more popular than, discriminant analysis, and perhaps in large part due to its recent availability in statistical packages (Montgomery *et al.*, 1987) <sup>[16]</sup>.

The intercept and the coefficients can be estimated using discriminant analysis and logistic regression. Both techniques operate in a similar manner but each has its underlying assumptions. Discriminant analysis programs usually provide a score. The ability to correctly identify (classify) those who have defaulted and those who have not defaulted is then evaluated at various cut-off values of the score. Most data analysis softwares will also provide the probabilities of group membership (e.g. default versus no default) on request. With logistic regression, the model is taken a step further by most data analysis softwares and the actual probabilities of the event occurring (e.g. default) are calculated. The classification capacity is appreciated. In this regard, the alternative methods utilize similar information. There are basic differences in the statistical assumptions of the two approaches, however. With discriminant analysis, the assumptions are that the independent variables are normally distributed, with equal variance within each group, and that the interrelationships (correlations) among the variables within each group are the same (i.e. equal covariance matrices) (Morrison D. G., 1969) <sup>[17]</sup>. These assumptions are usually broken when some or all of the independent variables are attribute type variables (i.e. present or absent). This is particularly important when the objective is to estimate the magnitude of the effects (coefficients) of the predictor variables. Using discriminant

analysis, the sign of the coefficients will always be correct, B's equal to zero will be estimated accurately, but estimates of non-zero B's may be biased (Halperin *et al.*, 1971). When the objective is overall prediction or classification as in this study, these assumptions are less constraining, and both methods should provide essentially the same model. The assumptions underlying the discriminant model need not be met when using logistic regression models. Experience also indicates that the coefficients of the logistic model are usually slightly closer to the null value of one, than those derived from discriminant analysis (Kleinbaum *et al.*, 1982). Several authors have formally compared the two techniques. For example, Halperin *et al* (1971) obtained results with from none to several attribute type predictor variables, and noted only small differences in the classification ability between the two methods. Kleinbaum *et al* (1982) <sup>[14]</sup> compared the classification ability of logistic regression and discriminant analysis using a data set which met the assumptions of discriminant analysis and noted that the logistic model was slightly superior. Montgomery *et al.* (1987) used two data sets from the prediction of coliform mastitis in dairy cows and suggest that logistic regression made few classification errors than discriminant analysis. Balogun O. S. *et al.* (2015) <sup>[5]</sup> used both approaches to classify the mode of delivery of an expectant mother and find that they estimated almost the same statistical coefficients and the overall classification for both was good. Press and Wilson (1978) <sup>[18]</sup> concluded that each technique served a unique function: discriminant analysis was useful for classification of observations into one of two or more populations, whereas logistic regression was useful for relating a qualitative (binary) dependent variable to one or more independent variables by a logistic distribution functional form of P (as noted above). The general conclusions are that if the assumptions of normality and equal variance/covariance matrices are met, then discriminant analysis estimators are preferred for the task of classification. If, however, the assumptions are not met, then

the logistic regression estimates are preferred for either application (Press and Wilson, 1978) [18]. The objective of this paper is to compare the two methods of analysis for classifying borrowers into one of two groups.

Eddokpayi *et al* (2013) [7] compared the two methods in classifying and assessing the relative importance of the fruit form characteristics, but concluded that the two methods were of nearly equal value but logistic regression would be preferable whenever the normality assumption are violated. Balogun *et al* (2014) [6] compared the two methods in classifying and assessing drug offender characteristics, but concluded that the two methods gives closely value but logistic regression would be preferable whenever the normality assumption are violated. In choosing between the two methods, the study applied the following criterion, the prediction of group membership and the assessment of its success i.e. determine which between the two methods provides a higher accuracy in classifying the mode of delivery of an expectant mother. Determine which variables appears significant in classifying the dependent variable by inspection of the coefficients and testing the assumption of normality and equal covariance required for the validity of the discriminant analysis. The outcome will not only complement the breeders' current practices but will also assist the research scientists to make appropriate choice in their application of these two techniques. In an earlier study, discriminant analysis was used to select useful variables and build a model for the prediction of default in the loan portfolio of MFIs. In this study we compared the results of discriminant analysis to those of logistic regression using the same data. Therefore, in this paper we examine the prediction of default in the loan portfolio of MFIs using logistic regression and discriminant analysis. Borrowers in the portfolio those who have never missed a loan repayment are in the no default group while those who have a missed loan repayments are in the default group. We use predictive variables to examine how the outcome of the predictions re different with the two statistical methods.

## 2- Methodology

In this section the method of data collection is presented followed by the methods of data analysis.

### 2.1. Method of Data Collection and source of Data

The data used in this study is collected from microfinance institutions and their borrowers with the collaboration of credit officers. The survey is carried out with the aid of questionnaires and direct interviews with loan officers and borrowers of some selected MFIs in urban areas of Cameroon. It is carried out in the selected towns of Douala, Yaounde, Bamenda and Limbe. The sample used in this thesis was made up of 723 borrowers. The data was collected from microfinance borrowers and the administration of MFIs. The collected data refers to the operations of credit granted to borrowers of microfinance institutions in Cameroon in the period 2016 to 2017. All the credit granted in that period was analysed with the purpose of verifying the probability of default of the borrowers.

The customers selected through the process of random sampling had their business records obtained from the administration of the MFIs and the customers themselves. We identified the personal characteristics of the customer and economic-financial data related to the business or economic activity of the customers.

## 2.2 Methods of data Analysis

In this sub section we present the two methods of data analysis which are respectively logistic regression and discriminant analysis.

### 2.2.1. Logistic Regression

The logit analysis or logistic regression consist of a statistical technique used in the separation of two groups which seek to obtain the probability that an observation belongs to a certain group, according to the behaviour of the independent variable  $s$  (Hosmer D. W., 2000) [13]. It is used commonly for analysis of data with binary or dichotomous answers and it consist of listing though model the answer variable (binary dependent variable) with factors that influence or not the probability of occurrence of certain events (independent variables). In the logistic regression the dependent variable since it possess a non-metrical feature is inserted through the use of a dummy variable (dichotomous or binary) that has a 0 value to indicates the absence of an attribute and to indicate the presence of an attribute (Agreti A., 2002) [1]. In the scope of the application of credit risk the technique of logistic regression is used for the evaluation of insolvency of a certain group of consumers in situations related to credit concessions assuming that the insolvency probability is distributed logistically with a binomial result 0 or 1. Supposing a dependent event in which the Y variable is a binary variable that assumes 0 or 1 values and independent variables  $X_1, X_2, \dots, X_k$ , the function of logistic distribution according to Gujarati (2000) is given by:

$$P_i = \Pr[Y_i = 1 | X = X_i] = \frac{1}{1 + e^{-Z}}$$

Where  $P_i$  represents the probability of an event occurring

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i$$

In analysing the meaning of the function of logistic regression in the credit risk context, there is the dependent variable which is a situation of the customer's prompt payment that will acquire 0 or 1 values depending on whether the data is coming from a prompt paying or insolvent customer or a solvent and insolvent company respectively; the independent variables represent the factors that supposedly influence insolvency as for example personal, economic or financial data of the individual customers, financial indexes in the case of companies. The probability of the customer's insolvency or of insolvency of the company is given by  $P_i$  which the conditional probability of Y is taking on a value 1. The estimated coefficient  $\beta_i$  represents measures of the variations in the proportions of the probabilities.

### 2.2.2 Linear Discriminate Analysis

The aim of linear discriminate analysis (herein after 'LDA') is to classify a heterogeneous subsets and further the decision process on these subsets. We can assume that for each applicant there are a specific number of explanatory variables available. The idea is to look for such a linear combination of explanatory variables, which separates most subsets from each other. In a simple case of two subsets, the goal is to find the linear combination of explanatory

variables which leaves the maximum distance between means of the two subsets. In a general case we consider the distributions  $p(x|G)$  and  $p(x|B)$  which are multivariate normal distributions with the common variance are. Then the equation above reduces to

$$A_G = \{x | \sum w_i x_i > c\}$$

As it follows from the econometrics theory, here  $x_i$  are explanatory variables,  $w_i$  are associated coefficients (weights) in the linear combination of explanatory variables. If one takes  $s(x) = \sum w_i x_i$  then it is possible to discriminate according to this score and thus to reduce the problem to only one dimension. The discriminate analysis was introduced by Fisher (1936) who searched for the best way to separate two groups using linear combination of variables [1]. Eisenbeis (1977) [8] criticised this method by stating that the rule is optimal only for a small class of distributions. However, Hand and Henley (1970) claim that ‘‘if the variables follow a multivariate ellipsoidal distribution (of which the normal distribution is a special case), then the linear discriminate rule is optimal’’.

Other critics state that there is a selection bias due to the fact that a learning sample for a credit scoring system is made of applicants to whom credit has been granted. That means that results are biased when applied to the whole population. Eisenbeis (1977) [8] also saw problems in the definition of the bad and good groups in the case when no clear boundary is between them and under the assumption that the covariance matrices of the two distributions are equal. In this case the use of quadratic discriminate analysis instead of the linear case is appropriate. Problems also arise when one wants to test for the significance of individual variables as one does not have the assumption of normality and therefore cannot perform statistical inference [2]. (Altman (1968) [2], who was the first to apply discriminate analysis, constructed the so-called z-score which is a linear corporate credit granting problem [3]. He found the model to be extremely accurate in correctly predicting bankruptcy. As we have mentioned, the advantages of the LDA method are that it is simple, it can be very easily estimated and it actually works very well; it is often used by financial institutions for credit scoring purposes. The disadvantage is that LDA requires normally distributed data but the credit data are often non-normal (and categorised).

### 2.3. Definition of Variables

The selection of variables is from the evaluation of the socio-economic and demographic environment of Microfinance institutions in Cameroon. It was equally based on the information available. A major element in the selection of variables was the information provided by credit officers. From the theoretical framework developed and the availability of information we have identified and collected information on the variables in the table below (some of the variables are defined in the tables).

<sup>1</sup> Fisher (1936) suggested (under assumption of common sample variance) looking for the linear combination of explanatory variables which leaves the maximum distance between means of the two classes.

<sup>2</sup> Many of these issues are addressed in the review by Rosenberg and Gleit (1994)

<sup>3</sup> The variables used are Sales/Total Assets (TA), working capital/TA, Retained Earnings/TA, Earnings before Interest and Taxation/TA, Market value of equity/Book Value of total debt.

**Table 1:** Variables in the models

Variable	Definition
Gender	Sex of the borrower
Marital	Marital Status of borrower
Education	Educational level of borrower
Bustype	The type of business carried
Exincome	The extra income of borrower
Exloan	Extra loans received by the borrower
Age	The age of the borrower
Numdepend	The number of depends under the responsibility of the borrower
Busrev	The amount of business revenue earned by the borrower
Misrepay	The number of repayments missed by the borrower
Othdebts	Other debts contracted by the borrower
Requestamt	The amount of loan requested by the borrower
Amntacg	The amount of loan actually granted to the borrower
Cotype	The type of collateral offered by the borrower
Numloans	Number of loans granted by institution to borrower
Reduration	The period of time required to refund the loan
Exloanoff	The number of years of experience of credit officer

Source: Author

The dependent variable in the two models is ‘‘missed loan repayment’’. It is a dummy variable in which 1 if the borrower did not miss any loan repayment and 0 if the borrower missed a loan repayment. After review of literature (Schreiner [4], 2004), the independent variables are arranged in three categories according to the factors affecting loan default. Characteristics of the borrower, characteristics of the loan and those related to the experience of the credit officer. The characteristics of the borrower provide insight into the likelihood of repayment. Knowing about the borrower’s external conditions (socio-economic) is an important way of knowing what factors could reduce their willingness to repay. The characteristic of the loan is an important factor because of how the demanding contractual conditions of MFIs can induce or discourage borrowers to respect their obligations. The credit officer plays an important role in identifying a bad borrower by evaluating their credibility prior to granting the loan. The inclusion of all variables would make the model unnecessarily large and scare customers when confronted with the required number of questions. There are several approaches in selecting independent variables proposed by authors. Hand and Henley (1997) [11] describe three ways of selecting variables. First, expert knowledge is used to select the right variables. Secondly statistical procedures such as forward and backward selection based on  $R^2$  can be implemented.

A combination of the forward and backward approaches the stepwise approach also exists. The techniques Logistic regression and Discriminant Analysis were implemented in classification groups of default or no default in the loan portfolio.

### 3. Estimated Results and Discussions

This section begins by presenting and discussing the descriptive statistics of the data collected then is continues with the results of the logistic regression and discriminant analysis.

<sup>4</sup> Schreiner classified the explanatory variables affecting default into three categories the characteristics of the borrower, the characteristics of the loan and variables related to the experience of credit officer.

### 3.1 Statistical Description of Variables

From table 2 below 25,6% of the female borrowers did not default as against 21,8% who defaulted indicating that males in our sample are more likely to default than females thus gender affects loan default. The results are different from those of some authors who show that women are better credit risks than men especially in the African context (Vigano, 1993). 37.2% of the respondents who had an

educational level more than primary school did not default suggesting that borrowers who are more educated are less likely to default than those who are more educated. This confirms the importance of education as a possible determinant of default. Most of the borrowers in our sample are married (61.1%) and 19% defaulted as against 42.1% who did not default. Seemingly, being unmarried increases credit risk.

**Table 2:** Descriptive statistics of variables

Variables		default	No default	Significance	S.E
Gender	Female	21.5%	25.6%	0.630	0.006
	male	28.4%	24.4%		
Educational level	Less than primary	27.3%	14.9%	0.002	0.037
	More than primary	20.7%	37.2%		
Extra income	Otherwise	30.6%	3.3%	0.000	0.073
	yes	17.4%	48.8%		
Age	18 to 25 years	8.6%	1.3%	0.000	0.058
	26 to 35 years	15.7%	1.7%		
	36 to 45 years	12.4%	20.7%		
	46 to 55 years	9.9%	21.5%		
	More than 55 years	0.7%	7.6%		
Marital Status	Unmarried	28.9%	9.9%	0.000	0.082
	married	19.0%	42.1%		
Type of activity	Agriculture	0.8%	5.0%	0.056	0.072
	Business	45.1%	49.1%		
Loan term period	Less than 1 year	37.2%	39.7%	0.856	0.091
	More than 1 year	10.7%	12.4%		
Extra loan	No	34.5%	23.5%	0.001	0.087
	Yes	26.8%	55.6%		
Number of dependent	1 to 2 people	16.5%	8.3%	0.000	0.071
	2 to 4 people	14%	17.4%		
	more than 4 people	9.1%	26.4%		
Revenue of borrower	Less than 500 000	18.2%	6.6%	0.014	0.020
	500 001 to 1 000 000	13.2%	19%		
	1 000 001 to 1 500 000	8.3%	18.2%		
	1 500 001 to 2 000 000	6.4%	6.8%		
	More than 2 000 000	1.5%	1.9%		
Other debts	No	12%	47%	0.000	0.063
	Yes	37.6%	3.4%		
Requested amount of lo	Less than 500 000	23.1%	11.6%	0.010	0.078
	001 to 2 000 000	14.9%	19%		
	2000 001 to 5 000 000	8.3%	14.9%		
	More than 5 000 000	1.7%	6.6%		
Amount granted	Less than 500 000	29.8%	24.8%	0.013	0.081
	001 to 2 000 000	11.6%	19%		
	2 000 001 to 5 000 000	3.3%	8.3%		
	More than 5 000 000				
Type of Collateral	Land	9.9%	24.8%	0.004	0.079
	Guarantor	19.8%	14.9%		
	Salary	9.6%	10.2%		
	Savings	2.7%	0.6%		
	house	5%	2.5%		
Number of loans received	1 only	35.5%	22.3%	0.000	0.085
	More than 1	12.4%	29.8%		
Repayment arrears	No	13.2%	40.5%	0.000	0.079
	Yes	34.7%	11.6%		
Requested duration	1 to 6 month	1.7%	9.9%	0.002	0.083
	7 to 12 months	18.2%	6.6%		
	13 to 18 months	14.2%	15.5%		
	19 to 24 months	8.3%	15.7%		
	More than 24 months	6%	4%		
Loan officer experience	Less than 1 year	11.6%	1.7%	0.000	0.050
	to 2 years	19%	9.9%		
	2 to 5 years	12.4%	17.4%		
	more than 5 years	4%	19.1%		

Source: Author

### 3.2 The Predictive Power and Sensitivity Analysis of logistic regression

In this section we analyse the predictive power and analyse the sensitivity of the model. It verifies the ability of the logistic regression to foresee default in the loan portfolio of a Microfinance Institution.

**Table 3:** Predictive power of logistic regression model at cut off point of 0.5

Observed	Predicted		
	Default		Percentage correct
	default	no default	
No default	0	330	0
Default	0	312	100.0
Overall percentage			51.4

Source: Author

312/642 = 48.6% of the borrowers defaulted and 51.4% did not default. Since our strategy is to predict that the borrower will default using this in predicting default you would be correct 48.6% all the times. It also suggests that if we knew nothing about our variables and guessed that a borrower would miss a loan repayment we would be correct 48.6% of the time.

### 3.6 The Discriminant Analysis Results

The Eigen value of 3.882 shows that a high proportion of the variance of the dependent variable is explained by the discriminant function. The canonical correlation (0.892) shows a high degree of association between discriminant function and dependent variable. The square of the canonical coefficient (0.795) equally indicates that a high proportion of the variance are explained by the dependent variable.

**Table 4:** Adequacy of the discriminant analysis

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	3.882 <sup>a</sup>	100.0	100.0	.892

Source: Author

Wilk's Lambda of 0.205 shows that about 20.5% of the independent variables are explained by the discriminant function which is not significant and indicates that the group means differ. The model is therefore adequate for the prediction of good and bad borrowers.

**Table 5:** Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.205	149.836	21	.000

Source: Author from SPSS analysis

It is seen that the discriminant analysis effectively predicts default by separating good and bad borrowers but it does not meet the underlying assumptions.

### Comparison of the classification of logistic regression and discriminant analysis

It is verified from the classification table that the logistic regression obtained a good predictive accuracy in its classification reaching 83.2% accuracy. The model correctly classifies 306/330 = 92.7% of the borrowers as those who have not defaulted (good borrowers) which is the percentage of occurrences correctly predicted thus indicating how

sensitive the model is. It equally correctly classifies 228/312=73.1% as defaulted which is the percentage of non-occurrences correctly predicted indicating the specificity of the model. The overall predictive accuracy of the model is 81.6%.

**Table 6:** Classification of logistic regression model

Observed	Predicted		
	Default		Percentage Correct
	Default	No default	
Default	228	84	73.1
No default	24	306	92.7
Overall predictive accuracy			83.2

Source: Author

On the other hand the sensitivity of the discriminant analysis is 96.6% which means that 84.5% of the bad borrowers are correctly classified and there are few false negative results (type II error). The model is highly specific (92.1%) meaning that 92.1% of good borrowers are correctly classified and there are few false positive results (type I error).94.2% of the originally group cases are correctly classified. The discriminant analysis successfully separates the groups (bad and good borrowers maximally but it does not produce the probability of default unlike logistic regression.

### 4-Conclusion

There is no conclusive argument on which statistical technique is the best (Thomas, 2003). Any model used can cut operating cost by making the loan process simple, reduce potential loan losses and focus attention more on problem loans. It is in this light that this paper recommends focuses on the logistic regression technique since it has good statistical properties, no strict assumptions on data and shows the probability of default.

### References

1. Agreti A. Categorical Data Analysis. Second Edition. New York. John Wiley and Sons, 2002.
2. Altman EI. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. The Journal of Finance. 1968; 23:589-609
3. Anderson R. The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation. Oxford University Press. New York, 2007.
4. Anderson R. Credit Risk Management. Presentation at the annual Meeting of NCT-1994. Kansas, MO, 2003.
5. Balogun OS. "An Assessment of the performance of discriminant analysis and logistic regression methods in the Classification of Modes of Delivery of An expectant Mother" Mathematical theory and Modeling, 2015, 5(5).
6. Balogun OS. A Comparison of the performance of Discriminant Analysis and the Logistic Regression methods in Classification of Drug Offenders in Kwara State. International Journal of Advanced Research. 2014; 2(10):280-286.
7. Edokpayi AA, Agho C, Ezomo JE, Edosomwan OS, Ogiugo OG. A Comparison of the Classification Performance of Discriminant Analysis and the Logistic Regression Methods in Identification of Oil Palm fruit Forms. A Paper Presented at the Annual Conference of

- Nigerian Statistical Association. 11-13th, September, 2013, 20-26.
8. Eisenbeis RA. Problems in applying discriminant analysis in credit scoring models. *Journal of Banking & Finance*. 1977; 2:205-219.
  9. Fisher RA. The use of multiple Measurements in Taxonomic Problems. *Anal of Eugenic*. 1936; 7(1936):179-188
  10. Godfrey K. Statistics in practice: Comparing the means of several groups. *New Engl J Med*. 1985; 313:1450-1456
  11. Hand D, Henley W. Statistical Classification Methods in Consumer Credit Scoring: A Review. *Journal of Royal Statistical Society series (statistics in Society)*. 1997; 160(3):523-541.
  12. Hand D, Henley W. Can reject Inference Ever work? *Journal of Management Mathematics*. 1993; 5(1):45-55.
  13. Hosmer DW, Lemeshow S. *Application Logistics Regression*. New York. John Wiley and Sons, 2000.
  14. Kleinbaum DG, Kupper LL, Muller KE, Morgens-Tern H. *Epidemiologic Research: Principles and Quantitative Methods*. Toronto: Lifetime Learning, 1982, 461-470.
  15. Krieng Kitbumrungrat. "A comparison of Logistic Regression and Discriminant Analysis in Classification group for Breast Cancer". *International Journal of Computer Science and Network Security*, 2012, 12(5).
  16. Montgomery ME, White ME, Martin SW. A Comparison of Discriminant analysis and Logistic Regression for the prediction of Coliform Mastitis in dairy Cows: *Canadian Journal of Veterinary Research*. 1987; 51(4):495-498.
  17. Morrison DG. On the interpretation of discriminant analysis. *J Marketing Res*, 1969, 156-163.
  18. Press J, Wilson S. Choosing Between Logistic Regression and Discriminant Analysis: *Journal of the American Statistical Association*. 1978; 73(364):699-705.
  19. SAS Institute Inc. *SAS's user's guide: statistics*. (1982) edition. Cary, North Carolina: SAS Institute Inc. 1982:381-396.
  20. Schreiner M. Scoring arrears at a microlender in Bolivia. *Journal of Microfinance*. 2004; 6:65-88.
  21. Schreiner M. Scoring Drop out at a Microlender in Bolivia, *Savings and Development*. 2003; 2(7):101-118
  22. SPSS Inc. *SPSS' user's guide*. New York: McGraw-Hill, 1983, 623-645.
  23. Thomas L. A survey of Credit and Behavioural Scoring; Forecasting Financial risk of lending to Consumers. *International Journal of Forecasting*, 2000, 16(4)
  24. Verter JI. Estimation of the multivariate logistic risk function: A comparison of the discriminant function and maximum likelihood approaches. *J Chron Dis*. 1971; 24:125-158.
  25. White ME, Glickman LT, Barnes-Pallesen FD, Pearson EG, Montgomery ME, *et al*. Discriminant analysis of the clinical indicants for bovine coliform mastitis. *Cornell Vet*. 1986; 76:335-341.
  26. White ME, Glickman LT, Barnes Pallesen FD, Stemes III, Dins-More P, Powers MS, *et al*. Accuracy of a discriminant analysis model for prediction of coliform mastitis in dairy cows and a comparison with clinical prediction. *Cornell Vet*. 1986; 76:342-347.