Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi/ Journal of The Institute of Natural & Applied Sciences 20 (1-2) 34-46, 2015

Araştırma Makalesi/Research Article

Comparison of Discriminant Analysis and Logistic Regression Analysis: An Application on Caesarean Births and Natural Births Data

Qais Mustafa ABDULQADER*

Zakho Technical Institute
Duhok Polytechnic University
Duhok, Iraq
*Qais.mustafa@dpu.edu.krd

Abstract: The discriminant analysis DA and the logistic regression analysis LRA are two statistical techniques used for analyzing data and predicting group membership from a set of predictors. Many applications have been done in this area. In this paper, we have been focus for the comparison of the two statistical techniques through applying on real data and exactly on caesarean births and natural births data. The comparison was depending on two statistical criteria; apparent error rate AER and apparent correct classification rate ACCR and performing stepwise procedure.
The results of the analysis showed that the performance of both techniques gave high ability in discriminating the kind of birth, whereas DA slightly exceeds LRA in the apparent correct classification rate and performed better than LRA in the births data. On the other hand, the results of DA showed that out of ten predicted variables, seven predictors exhibited strong evidence in classifying and discriminating the kind of birth, while the results of LRA showed that six predicted variables out of ten predictors have contributed significantly to discriminate the kind of birth. The suitable model for both techniques has been estimated depending on the selected predictors.

Keywords: Apparent Error Rate, Apparent Correct Classification Rate, Caesarean Births, Discriminant Analysis, Logistic Regression Analysis, Natural Births

Diskriminant Analizi Karşılaştırılması ve Lojistik Regresyon Analizi: Sezaryen Doğum ve Doğal Doğum Veri Üzerine Bir Uygulama

Özet: Diskriminant analizi (DA) ve lojistik regresyon analizi (LRA) bir dizi belirleyiciden veri analizi ve grup üyeliğinin tahmininde kullanılan iki istatistiksel tekniktir. Bu alanda birçok uygulama yapılmaktadır. Bu makalede, sezaryen doğum ve doğal doğum gerçek verileri uygulanarak iki istatistiksel tekniğin karşılaştırılmasına odaklanılmıştır. Karşılaştırmada iki istatistiksel kriterlere bağlı kalınmıştır; belirgin hata oranı (BHO) ve görünür doğru sınıflama oranı (GDSO) ve aşamalı işlemler gerçekleştirilmiştir.
Analiz sonuçları, her iki tekniğin performansının doğum türünü ayırt etmede yüksek yeteneğe sahip olduğunu gösterirken, belirgin doğru sınıflandırma oranında DA, hafifçe LRA'yı aşmış ve doğum verilerinde LRA'dan daha iyi bir performans göstermiştir. Öte yandan, DA sonuçları, on tahmini değişken içerisinde yedi tanesinin sınıflandırılma ve ayırt etmede güçlü kanıtlar sergilemiştir olduğunu gösterirken, LRA sonuçları, on tahmini değişken içerisinde altı tanesinin doğum türünün ayırt edilmesine katkıda bulunduğunu göstermiştir. Her iki teknik için uygun bir model, seçilen belirleyicilere bağlı olarak tahmin edilmiştir.

Anahtar kelimeler: Ayırma Analizi, Lojistik Regresyon Analizi, Sezaryen doğum, Doğal Doğum, Görünür Hata Oranı, Görünür Doğru Sınıflandırma Oranı
MSC 2010: 62H30, 62H99

1. Introduction

The DA is a multivariate statistical technique used to estimate the linear relationship between a dependent variable having two or more categories and linear combinations of more independent variables. In other words DA concerned with separating distinct sets of objects or observations and with allocating new objects to previously defined groups (defined by a categorical variable), Johnson and Wichern (2007). The LRA analyze the relationship between multiple independent variables and a single binary dependent variable, categorical variable with two categories, Sweet and Martin (2003).

Many applications have been done in the area of LRA and DA. Pohar et al. (2004) compared the robustness of the DA and LRA methods towards categorization and non-normality of explanatory variables in a closely controlled way. They showed that the results of DA and LRA are close whenever the normality assumptions are not too badly violated. Panagiotakos (2006) compared between logistic regression and linear discriminant analysis for the prediction of categorical health outcomes. He concluded that, logistic regression resulted in the same model as did discriminant analysis. Charlo (2010) presented an empirical study about risk analysis using several technologies (DA, LRA and an artificial intelligence technology AIT). The percentage of error when using an AIT concluded that these intelligent systems are a good support in decision-making for risk analysts from banking entities. Zhu and Li (2010) applied the DA and LRA model in credit risk for China's listed companies. The results of empirical research showed that LRA model is superior to DA model. Amin et al. (2011) captured the determinants of consumer preference for genetically modified palm oil that has less saturated fat using DA and binary logistic regression BLR. Results of the study implied the importance of credible and effective dissemination of consumer information by the relevant authorities in the country.

Recently, the DA and also LRA have been used in medical studies. Maiprasert and Kitbumrungrat (2012) have made comparison between multinomial logistic regression analysis MLRA and DA in predicting the stage of breast cancer. The results of an analysis revealed that the MLRA was capable of 55.50 percent correctly predicting in overall, which was more correct than the analysis done by DA, giving a 54.10 percent correct prediction in overall. Zandkarimi et al. (2013) have identified the determinants for diabetes in people with pre-diabetes employing two advanced statistical methods of LRA and DA. The results showed that the predictive power of LRA and DA were 0.884 and 0.80 respectively concluding that the LRA is more powerful in the separation of patients from pre-diabetic. Shaheen (2014) have been focused in his paper for the comparison between three forms for classification data belongs two groups when the response variable with two categories only. The results showed that the probability form of the LRA has minimum probability of misclassification through the application on the data of two types of leukemia.

Very recently, Balogun et al. (2015) compared DA and the LRA model in predicting mode of delivery of an expectant mother, natural birth and caesarian section. They showed that both methods were of nearly equal value 64.7% and 65.8%, and almost selected the same set of variables and also showed that mother's weight is very significant to identifying expectant mother's mode of delivery however, given the failure rate to meet the underlying assumptions of DA, LRA is preferable. Gjonej et al. (2015) have made a descriptive study for the reasons of rising trend of caesarean births rate year after year. The results showed that previous caesarean birth and multiple pregnancies represent a growing trend.

In this paper, two different techniques were considered and applied on births data. In the first technique, the births data is modeled using DA. In second technique the births data is modeled using LRA. Then some statistical criteria were computed for each technique and have used for evaluation and comparison. The remainder of this paper is structured as follows: In next section we review the objectives of this paper and gives the explanation of DA and LRA. Section 3 deals with application and main results of both techniques and gives discussion. Finally in section 4 conclusions are presented.

2. Objectives and the Explanation of DA and LRA

2.1. Objectives

The first objective of this paper is to compare the DA technique with LRA technique so as to investigate whether these two techniques of analysis have the same results or not. The second objective is to obtain the suitable model for diagnosis, classification and for discriminating between the type of birth (i.e.: caesarian births and natural births). The procedure of comparison and choosing the suitable model is depending on some statistical criteria.

2.2. The Explanation of DA and LRA

2.2.1. DA

DA is a statistical multivariate technique used in many different fields and it involves a discriminant variety and represents a linear combination of two or more predictors that discriminate between the objects in the groups defined a priori, Joseph et al. (2010).

2.2.2. Assumptions of DA

We can summarize the assumptions of the DA as: Homogeneous within – group variances, multivariate normality within group, linearity among all pairs of variables, no multicollinearity and prior probabilities, Joseph et al. (2010).

2.2.3. Linear Discriminant Function Model

The discriminant function takes the form of the linear equation:

$$Z_{jk} = \alpha + W_1 X_{1k} + W_2 X_{2k} + \cdots + W_n X_{nk}, \tag{2.1}$$

Where Z_{jk} is discriminant Z score of the discriminant function j for object k, α the intercept, W_i represent the discriminant weight for predictor i, and X_{ik} is the predictor i for object k, Joseph et al. (2010). The probability that a case with a discriminant score of Z belongs to group i is estimated by the following equation:

$$\pi(G_i|D) = \frac{\pi(D|G_i)\pi(G_i)}{\sum_{j=1}^{n} \pi(D|G_j)\pi(G_j)}, \tag{2.2}$$

Where the prior represented by $\pi(G_i)$ is an estimate of the likelihood that a case belongs to a certain group. The objects are classified into one or the other group on the basis of the obtained Z score, whether it is higher or lower than the predefined cut off value, Memic (2015).

2.3. LRA

LRA is a special case of linear regression analysis used when the dependent is dichotomous/ordinal (ordered categories) not continuous and the predictor variables are metric or nonmetric variables, Joseph et al. (2010). It was first proposed in the 1970s as an alternative technique to overcome limitations of ordinary least squares OLS regression in handling dichotomous outcomes, Peng and So (2002). The goal is to predict a particular category for one or more predictor variables, which may or may not be continuous. Instead of OLS regression, maximum likelihood estimation MLE is used in LRA. In addition, the LRA allows us to predict the likelihood of a binary outcome based on many variables, including other binary variables, Anthony (2011). As a result, the statistics of interest are Wald-Chi-square values instead of F or t-values. Fortunately, the traditional method of model comparison and hypothesis testing is unchanged, Hosmer and Lemeshow (2000).

2.3.1. Assumptions of LRA

With LRA the assumptions are that a LRA exists between the probabilities of group memberships and a linear function of the predictor variables. It is also assumed that observations are independent. Moreover, in the LRA context, the more unequal the numbers in the categories, the more cases are needed. Add to all this the problem of missing data because of list-wise deletion, and the desirability of having enough cases to cross-validate results on a holdout sample, and it becomes painfully clear that LRA typically requires cases in the hundreds to guarantee trustworthy results. Additionally, there is no formal requirement for multivariate normality, homoscedasticity, or linearity of the predictor variables within each category of the dependent variable. We refer the reader to, Bewick et al. (2005) and Peng et al. (2002) for more details.

2.3.2. LRA Model

LRA is suitable for studying the relation between a categorical or qualitative outcome variable and one or more predictor variables. In case of one predictor X and one dichotomous outcome variable Y, the logistic model predicts the logit of Y from X. The logit is a natural logarithm of odds of Y. The simple formula of LRA model can be written as the following, Peng and So (2002), Hosmer and Lemeshow (2000).

$$\ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta x. \tag{2.3}$$

Hence, $\pi(x) = E(Y|X) = p(Y = \text{outcome of interest}|X = x) = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$, (2.4)

Where π is the probability of the outcome of interest, α is the Y intercept, and β is the slope parameter. X can be categorical or continuous variable, and Y is always categorical. The simple model can be extended to multiple logistic regression as follows:

$$\ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k. \tag{2.5}$$

Therefore,

$$\pi(x) = E(Y|X) = p(Y = \text{outcome of interest}|X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{e^{\alpha+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k}}{1+e^{\alpha+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k}}, \tag{2.6}$$

Where π is the probability of the event, α is the Y intercept, β_s are slope parameters, and X`s are a set of predictors. α and β_s are estimated by the MLE method.

2.4. The Performance Criteria

The most important think when building a classification rule is to correctly classify as many future units as possible. Several criteria are available for evaluate a set of classification rule and one of the simplest criteria is error rate or misclassification rate, Elhabil and Eljazzar (2014). A simple estimate of the error rate can be obtained by trying out the classification procedure on the same data set that has been used to compute the classification functions. This method is commonly referred to as re-substitution, Rencher (2002). For two groups, among the n_1 observations in G_1, n_{11} are correctly classified into G_1, and n_{12} are misclassified into G_2, where $n_1 = n_{11} + n_{12}$. Similarly, of the n_2 observations in G_2, n_{21} are misclassified into G_1, and n_{22} are correctly classified into G_2, where $n_2 = n_{21} + n_{22}$. Thus, the apparent error rate AER can be presented as:

$$\mathrm{AER} = \frac{n_{12} + n_{21}}{n_1 + n_2} = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}}. \tag{2.7}$$

Similarly, we can present apparent correct classification rate ACCR as:

$$\mathrm{ACCR} = \frac{n_{11} + n_{22}}{n_1 + n_2}. \tag{2.8}$$

3. Application and Main Results

3.1.Application on Real Data

The application concerning the comparison of the DA and LRA was performed using the births data and working with SPSS statistical package program. The data were collected by the statistics unit in Maternity Hospital in Zakho district, Kurdistan Region of Iraq. The classification task consists of predicting whether a birth would test positive for caesarean. The births data were labeled, such that we put 1 for caesarean birth and 0 for non- caesarean birth or natural birth. There are 10 predictor variables for 230 births, and among them, 140 births tested positive for caesarean. Table 1 presents the details about the frequency and percentage distribution of the groups:

Table 1. The Class Distribution of the Births Data

Class Name	Class Size	Class Distribution
Caesarean Birth or Positive	140	60.87%
Natural Birth or Negative	90	39.13%

As shown in table 1, the sample size of the data is 230 observations and the data set were divided in to two groups such that, the first group with n_1=140 which represents the 60.87% of observations, and the second group with n_2=90 which represents the 39.13% 0f observations. The dependent variable name is “Kind” which represents the kind of birth. The predictor variables are presented as the following: X_1 for gender, here we put 0 for male and 1 for female, X_2 for baby weight in Kg, X_3 for The number of weeks of pregnancy, X_4 for mother weight in Kg, X_5 for mother age in years, X_6 for the number of times pregnant, X_7 for previous caesarean, here we put 1 if there is a previous caesarean section operations and 0 otherwise, X_8 for breech baby, here we put 1 if the status of the fetus is breech and 0 otherwise, X_9 for mother risk, here we put 1 if there is a risk to the mother's life and 0 otherwise, X_{10} for mother wish, here we put 1 If the mother was wishes to caesarean and 0 otherwise. Table 2 shows the summary of statistical analysis which represents the mean and the standard deviation of the predictors.

Table 2. Summary of Statistical Analysis

Ser.	Predictor Variables	Mean	Standard Deviation
1	X_1	0.470	0.500
2	X_2	3.321	0.501
3	X_3	38.287	2.036
4	X_4	76.635	10.984
5	X_5	27.000	6.312
6	X_6	4.257	2.767
7	X_7	0.609	0.964
8	X_8	0.135	0.342
9	X_9	0.252	0.435
10	X_{10}	0.391	0.489

3.2.Application Using DA

The main assumptions of DA were tested. Kolmogorov-Smirnov test statistic used for testing normality of births data and had a value of 0.397 with p-value <0.001. Depending on the significance level =0.05, we conclude that the data are not multivariate normally distributed. This conclusion was expected because we are dealing with dichotomous or binary variable, but the sample size exceeds 30 observations and depending on the Central Limit Theory, we can say that the data follow the normal distribution. Box's M test used here to test the assumption of equality of covariance matrices and had a value of 5.45 with p-value 0.02 indicates that the data do not differ significantly from multivariate normal. Concerning the assumption of multicollinearity, high correlations should not be present among variables of interest. For this purpose the Variance Inflation Factor VIF test is used here, and a value of VIF >10 indicates multicollinearity is present and the assumption is violated, Berenson et al. (2012). Table 3 shows the VIF values of all predictors.

Table 3. VIF Values of Predictor Variables

Ser.	Predictor Variables	VIF Values
1	X_1	1.066
2	X_2	1.541
3	X_3	1.465
4	X_4	1.134
5	X_5	3.087
6	X_6	3.155
7	X_7	1.191
8	X_8	1.350
9	X_9	1.309
10	X_{10}	1.121

Observe that all the VIF values in table 3 are relatively small, ranging from a high of X_6=3.155 for the number of times pregnant to a low of X_1=1.066 for gender. Thus, on the basis of the VIF criteria, we conclude that all VIF values are less than 10, there is no evidence of multicollinearity among the set of predictor variables or predictors. This means one can proceed with the analysis.

Before creating the discriminant model, Wilks' Lambda statistic with F-test are performed here for testing the equality of group means in order to determine the variables which significantly contribute to the differentiation of groups and also to measure each predictor`s potential. It is equal to the proportion of the total variance in the discriminant scores not explained by differences among the groups. Smaller values of Wilks' Lambda indicate greater discriminatory ability of the function, Rencher (2002). As we mentioned in section (3.1), that we have 10 predictors, the stepwise method can be useful by automatically selecting the best variables to use in the model. Seven most important predictors were selected respectively X_4, X_{10}, X_7, X_8, X_2, X_3 and X_6 and others were removed from the analysis because they were not significance. Table 4 presents the selection of discriminating variables depending on stepwise method and performing Wilks' Lambda statistic.

Table 4. Selection of Discriminating Variables Depending on Stepwise Method

Step	Variables Entered	Wilks' Lambda	D.F.1	D.F.2	D.F.3	Exact F			
						Statistic	D.F.1	D.F.2	Sig.
1	X4	0.630	1	1	228	133.630	1	228	0.000
2	X10	0.584	2	1	228	80.872	2	227	0.000
3	X7	0.542	3	1	228	63.533	3	226	0.000
4	X8	0.492	4	1	228	58.102	4	225	0.000
5	X2	0.480	5	1	228	48.591	5	224	0.000
6	X3	0.462	6	1	228	43.348	6	223	0.000
7	X6	0.450	7	1	228	38.726	7	222	0.000

Wilks' Lambda can also be used to measure of how well each function separates cases into births groups. From table 5 we can conclude that the discriminant function is significant and the corresponding function explains the group membership well.

Table 5. Wilks' Lambda Table

Wilks' Lambda	Chi-square	D.F.	Sig.
0.450	179.149	7	0.000

3.2.1. Estimating the Discriminant Model

Table 6 gives unstandardized canonical discriminant function coefficients which are used in the formula for making the classifications in DA, much as. The constant plus the sum of products of the unstandardized coefficients with the observations yields the discriminant scores.

Table 6. Unstandardized Canonical Discriminant Model Coefficients

Variables	Coefficients of Function
X_2	-0.807
X_3	0.166
X_4	-0.087
X_6	-0.081
X_7	0.427
X_8	1.114
X_{10}	0.792
(Constant)	2.619

That is, the estimated discriminant model is:

$$Z = 2.619 - 0.807(X_2) + 0.166(X_3) - 0.087(X_4) - 0.081(X_6) + 0.427(X_7) + 1.114(X_8) + 0.792(X_{10}) \quad (3.1)$$

3.2.2.Testing the Discriminatory Power of the Discriminant Model

The two performance criteria AER and ACCR as we referred in section 2.4 are performed here to evaluate the efficiency of the discriminatory model of the estimated function. Table 7 shows the final results of classification.

Table 7. The Final Classification Results

Kind	Predicted Group Membership		Total
	Natural Birth	Caesarean Birth	
Natural Birth	75	15	90
Caesarean Birth	15	125	140
AER	13%		Grand Total=230
ACCR	87%		

From table 7 we can see that 75 of 90 births from the first group were correctly classified and 125 of 140 births from the second group were correctly classified. We conclude that the DA was able to classify 200 cases of births out of 230 cases correctly. The AER was 13% and the ACCR was 87% indicating that the model has high ability on classification.

3.3.Application Using LRA

LRA was applied on the births data set to study the relationship between the dependent variable (i.e., two groups of births) and combination of predictors in order to find the most important predictors that discriminate the kind of birth. Table 8 presents the model fitting information showing the statistical significance of the final chi-square.

Table 8. Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2Log Likelihood	Chi-Square	D.F.	Sig.
Intercept Only	307.891	170.002	6	0.000
Final	137.889			

It is obvious from table 8 that -2log likelihood value of basic model including only the intercept was 307.891 and this value have reduced to 137.889 with the existence the combination of predictors in the model. The value of the Chi-square was 170.002 against the probability 0.000 telling us that the model is significant and for this reason, we reject the null hypothesis and accept the alternative hypothesis which says that there is a real relationship between the predictors and the dependent variable.

There is a test of regression coefficient of the model using Likelihood Ratio Test LRT. This test evaluates the overall relationship between predictor and the dependent variable. By using the stepwise procedure, six most important predictors were selected respectively X_{10}, X_8, X_4, X_2, X_3 and X_7 and others were removed from the analysis because their contributions into the model did not play significant role to discriminate the kind of birth. Table 9 shows the result of the tests.

Table 9. Result of LRT

Effect	Model Fitting Criteria	LRT		
	-2Log Likelihood Of Reduced Model	Chi-Square	D.F.	Sig.
Intercept	140.775	2.886	1	0.089
X_{10}	153.737	15.848	1	0.000
X_8	157.110	19.221	1	0.000
X_4	211.140	73.251	1	0.000
X_2	147.243	9.354	1	0.002
X_3	143.450	5.561	1	0.018
X_7	158.439	20.550	1	0.000

3.3.1. Estimating the LRA Coefficients

Estimation of the model parameters obtained by using the MLE method with Wald statistic for the final model are presented in the table 10 which gives the results of fitting the LRA model to births data and presenting coefficients which are used in the formula for making the classifications in LRA, much as. The constant plus the sum of products of the coefficients with the observations yields the discriminant scores.

Table 10. Results of Fitting the LRA Model to Births Data

Variable	β	Std. Error	Wald	D.F.	Sig.	Exp.(B)
Intercept	-8.221	4.599	3.195	1	0.074	
X_{10}	-1.804	0.408	14.103	1	0.000	0.165
X_8	-3.327	0.936	12.626	1	0.000	0.036
X_4	0.174	0.027	42.137	1	0.000	1.190
X_2	1.822	0.640	8.116	1	0.004	6.184
X_3	-0.272	0.129	4.475	1	0.034	0.762
X_7	-1.548	0.436	12.583	1	0.000	0.213

That is, the estimated logistic regression model is:

$$\ln\left(\frac{\pi}{1-\pi}\right) = -8.221 + 1.822(X_2) - 0.272(X_3) + 0.174(X_4) - 1.548(X_7) - 3.227(X_8) - 1.804(X_{10}) \quad (3.2)$$

3.3.2. Testing the Discriminatory Power of the LRA Model

The two performance criteria AER and ACCR as we referred in section 2.4 are performed here to evaluate the efficiency of the LRA model of the estimated function. Table 11 shows the final results of classification.

Table 11. The Final Classification Results Using LRA model

Kind	Predicted Group Membership		Total
	Natural Birth	Caesarean Birth	
Natural Birth	76	14	90
Caesarean Birth	18	122	140
AER	13.9%		Grand Total=230
ACCR	86.1%		

From table 11 we can see that 76 of 90 births from the first group were correctly classified, and 122 of 140 births from the second group were correctly classified, we conclude that the LRA was able to classify 198 cases of births out of 230 cases correctly. The AER was 13.9% and the ACCR was 86.1% indicating that the model has high ability on classification.

3.4 Comparison and Discussion

The data was analyzed using DA and LRA. The results of DA showed that 87% of original cases were correctly classified to their respective group. The analysis also showed that the predictors;

X_4, X_{10}, X_7, X_8, X_2, X_3 and X_6 exhibited strong evidence in discriminating the kind of birth, while other predictors showed less contribution in explaining the variation between the two groups.

The results of LRA showed that 86.1% of the cases were correctly classified to their respective groups. The analysis also showed that the predictors; X_{10}, X_8, X_4, X_2, X_3 and X_7 exhibited strong evidence in discriminating the kind of birth. Comparing the results obtained from the above techniques indicate that the two techniques gave the best and strong ability in discriminating the kind of birth and they gave almost the same percentage of correct classification. In addition, the DA identified seven predictors while LRA identified six predictors responsible in discriminating the kind of birth. It should be noted that both techniques have chosen almost the same predictive variables.

4. Conclusion

In this paper, we have compared of two statistical techniques for classifying the births data which represents the DA depending on stepwise procedure and performing Wilks' Lambda statistic for selecting the most important predictors, and LRA depending on stepwise procedure and performing the LRT for selecting the most important predictors. The comparison was depending on two statistical criteria: AER and ACCR. From the results, we concluded that the performance of both analysis DA and LRA gave high ability of classification (ACCR for DA=87% and ACCR for LRA=86.1%). The results of LRA were very close to DA results, whereas DA slightly exceeds LRA in the ACCR, and performed better than LRA in the births data.

In addition, the DA analysis showed that the seven predictors; X_4, X_{10}, X_7, X_8, X_2, X_3 and X_6 which represents the predictor variables respectively (mother weight, mother wish, previous caesarean, breech baby, baby weight, number of weeks of pregnancy and number of times pregnant) exhibited strong evidence in discriminating the kind of birth and other predictors; X_1, X_5 and X_9 which represents the predictor variables respectively (gender, mother age and mother risk) were removed from the analysis because they were not significant statistically. The best model obtained by using DA through looking at the parameters of the discriminant model and its signs in equation (3.1) it is observed that the cases of caesarean births decreases with increasing the predictors; X_2, X_4 and X_6 which represents the negative relationship between the kind of birth and these predictors, and also the cases of caesarean births increases with increasing the predictors; X_3, X_7, X_8 and X_{10} which represents the positive relationship between the kind of birth and these predictors. On the other hand, the LRA analysis showed that the six predictors; X_{10}, X_8, X_4, X_2, X_3 and X_7 have contributed significantly to discriminate the kind of birth and the

remain predictors; X_1, X_5, X_9 and X_7 were removed from the analysis because they were unable to give a real or a positive contribution when discriminating the kind of birth. The best model obtained by using LRA through looking at the parameters of the discriminant model and its signs in equation (3.2) it is observed that the logit cases of caesarean births decreases with increasing the predictors; X_3, X_7, X_8 and X_{10} which represents the negative relationship between the logit of the kind of birth and these predictors, and also the logit cases of caesarean births increases with increasing the predictors; X_2 and X_4 which represents the positive relationship between the logit of the kind of birth and these predictors.
Comparing the results obtained from DA and LRA we conclude that the two techniques gave high percentage of correct classification, whereas DA slightly exceeds LRA in the ACCR, and identified predicted variables responsible in discriminating the kind of birth more than LRA.

Acknowledgements

The author would like to express gratitude to Mr. Idrees Haji the director of the statistics unit in Maternity Hospital in Zakho district for his kind help to get data.

References

Amin L., Othman J., Lip H., Goh and Josuff, K. (2011). Consumer preference for genetically modified (GM) food: The case of less saturated fat palm oil in Malaysia, African, Journal of Agricultural Research, Vol. 6, pp. 5212-5220.

Anthony, Denis (2011). *Statistics for health, life and social sciences*, Denis Anthony and Ventus publishing.

Balogun, O. S., Akingbade, T. J. and Oguntunde, P. E. (2015). An assessment of the performance of discriminant analysis and the logistic regression methods in classification of mode of delivery of an expectant mother, Mathematical Theory and Modeling , Vol. 5 , pp. 147-154.

Berenson, M. L., Levine, D. M. and Krehbiel, T. C. (2012). *Basic business statistics: Concepts and applications*, twelfth edition, Pearson education Inc.

Bewick, V., Cheek, L. and Ball, J. (2005). Statistics review 14: Logistic regression, Critical Care, Vol. 9, pp. 112-118.

Charlo, M. J. (2010).The most relevant variables to support risk analysis for loan decisions: An empirical study, Regional and Sectoral Economic Studies, Vol. 10, pp. 61-70.

Elhabil, A. and Eljazzar, M. (2014). A comparative study between linear discriminant analysis and multinomial logistic regression, An-Najah University Journal research. Vol. 28. pp. 1528-1548.

Gjonej, R., Keta, M., Zyberaj, F. and Smakaj, E. (2015). The reasons of rising trend of cesarean section rate year after year. A retrospective study, International Journal of Nursing and Midwifery. Vol. 7, pp. 9-15.

Hosmer, D. W. and Lemeshow, S. (2000). *Applied logistic regression*, Second edition, John Wiley And Sons, Inc.

Johnson, R. A. and Wichern, D. W. (2007). *Applied multivariate statistical analysis*, sixth edition, Pearson Prentice Hall.

Joseph, F. H. JR., William C. B., Barry, J. B. and Rolph, E. A. (2010). *Multivariate data analysis*, Seventh edition , Pearson Prentice Hall.

Maiprasert, D. and Kitbumrungrat, K. (2012). Comparison multinomial logistic regression and discriminant analysis in predicting the stage of breast cancer, International Journal of Computer Science and Network Security, Vol. 12, pp. 44-52.

Memic, D. (2015). Assessing credit default using logistic regression and multiple discriminant analysis: Empirical evidence from Bosnia and Herzegovina ,Interdisciplinary Description of Complex Systems, Vol.13, pp.128-153.

Panagiotakos, D. B. (2006). A comparison between logistic regression and linear discriminant analysis for the prediction of categorical health outcomes, International Journal of Statistical Sciences, Vol. 5, pp. 73-84.

Peng, C.Y. J. and So, T. S. H. (2002). Logistic regression analysis and reporting: A primer, Lawrence Erlbaum Associate Inc., Vol. 1, pp. 31-70.

Peng, C.Y. J., Lee, K. L. and Ingeresol, G. M. (2002). An introduction to logistic regression analysis and reporting, The Journal of Educational Research, Vol. 96, pp. 3-14.

Pohar, M., Blas, M. and Turk, S. (2004). Comparison of logistic regression and linear discriminant analysis: A simulation study, Metodoloski Zvezki, Vol. 1, pp. 143-161.

Rencher, A. C. (2002). *Methods of multivariate analysis*, Second edition, John Wiley and Sons, Inc.

Shaheen, H. I. (2014). Comparison between some of linear classification models with practical application, Journal of Economic and Management.Vol. 20, pp. 394-410.

Sweet, S. A. and Martin, K. G. (2003). *Data analysis with SPSS: A first course in applied statistics*, Second edition, Bosten: Ballyn and Bacon.

Zandkarimi, E., Safavi, A. A., Rezaei, M. and Rajabi, G. (2013). References comparison and discriminant analysis in identifying the determinants of type 2 diabetes among pre-diabetes of Kermanshah rural areas, Journal of Kermanshah University of Medical Sciences, Vol.17, pp. 300-308.

Zhu, K. L. and Li, J. j. (2010). Studies of discriminant analysis and logistic regression model application in credit risk for china’s listed companies, Management Science and Engineering, Vol.4, pp. 24-32.