International Journal of Bioassays
ISSN: 2278-778X
CODEN: IJBNHY
Open Access

Original Research Article

Evaluation Of Logistic Regression On Mode Of The Delivery Of Expectant Mothers

Balogun Oluwafemi Samson
Department of Statistics and Operations Research, Modibbo Adama University of Technology, P.M.B. 2076, Yola, Adamawa State, Nigeria.

Received for publication: February 12, 2015; **Accepted**: April 28, 2015

Abstract: The aim of the study is to obtain a Logistic Regression that can be used to classify the mode of delivery of pregnant women using some variables. Data from Health Records of 184 Pregnant women who delivered at the General Hospital, Wuse were used. The data consist of Mother's Weight, Height, Age and Baby's Weight, Baby's Gender and mode of delivery (Natural birth and Caesarian Section). In literature the author discovered that the discriminant analysis cannot handle mixed data (discrete and continuous data) effectively instead logistic regression better fit the data. The model correctly classified 65.8% original grouped cases with positive and negative predictive values 65.75% and 68.70% respectively, which are higher than values obtained under the discriminant analysis.

Key Words: Logistic Regression; Classification; Odds Ratio; Discriminant Analysis

Introduction

Child birth poses considerable risk to the lives of both mother and child particularly in situations where complication arises. Child birth is defined as the complete expulsion or extraction of a fetus from its mother. Child birth is preceded by a period known as the Gestation period. It has been of interest to researchers to know the mode of delivery a mother is likely to use. Under normal conditions, a mother is expected to give birth by natural birth otherwise known as safe delivery, but in certain cases complications may arise leading to the use of Caesarian section. Caesarian section poses considerable risk.

West/Central Africa accounts for more than 30% of global maternal deaths, and 162, 000 women died of pregnancy or childbirth related causes in 2005. The maternal mortality ration is substantially higher here than in any other region, at 1100 maternal deaths per 100, 000 live births. Furthermore, no discernible progress has been made in reducing the ratio since 1990. Of the 23 countries in the region with comparable estimates every country but Cape Verde has an MMR of at least 500, and a third of these countries have an MMR of 1, 000 or greater. Almost two thirds of maternal deaths in the region occur in the Democratic Republic of Congo, Niger and Nigeria, which together account for approximately 20 per cent of all maternal deaths world-wide. (UNICEF: Progress for Children Report 2008)

Several factors influence the high rate of maternal mortality in Nigeria, but the most common causes are lack of access to ante-natal care, inadequate access to skilled birth attendees, delays in the treatment of complications of pregnancy, poverty and harmful traditional practices.

Logistic regression is a mathematical modeling approach that can be used to describe the relationship between independent or predictor variables to dichotomous dependent variable (Kleinbaum *et al.,* 1998). The model has traditional been appealing due to its performance in classification, the potential to use its output as probabilistic estimates since they are in the range [0, 1], and interpretation of the coefficients in terms of the log-odds ratio. It is especially in biostatistical application where binary classification tasks occur frequently (Hastie *et al.,* 2001). Agresti (1996) asserted that relationships between the probability of success $\pi(x)$ and X are usually nonlinear rather than linear. Therefore a fixed change in X may have less impact when $\pi(x)$ is near 0 or 1 than $\pi(x)$ is in the middle of its range.

Logistic regression is often chosen if the predictor variables are a mixture of continuous and categorical variables and/or if they are not nicely distributed, that is, logistic regression makes no assumptions about the distribution of predictor variables. The logit analysis is usually employed if all the predictors are categorized while discriminant function analysis is used if all the predictors are continuous and nicely distributed (see Http://www.math.toronto.edu/mathnet).

Balogun *et al., 2014,* uses logistic regression to classify offenders into the two dichotomous dependent variable (peddlers or non-peddlers) in order to ease the work of the agency, National Drug Law Enforcement Agency (NDLEA), responsible for illicit drugs. The model correctly classified 95.4% original grouped cases with positive and negative predictive values 92.44% and 97.90% respectively, which are higher than values obtained under the discriminant analysis.

***Corresponding Author:**
Dr. Balogun Oluwafemi Samson,
Department of Statistics and Operations Research,
Modibbo Adama University of Technology,
P.M.B.2076, Yola, Adamawa State, Nigeria.

Since the aim of the study is to determine whether the mode of delivery of expectant Mothers can be classified as a natural birth/ Caesarian section on the basis of mother's weight, mother's height, mother's age, baby's age and baby's gender, therefore large quantity as never been reported in literature to have influenced the logistic regression model. However, some large quantity which could have been considered as outliers in the data collected was included/removed to determine the effect on the result of the study.

The logistic regression is employed for classification of objects into binary variable as

$$Y = \begin{cases} 1 & \textit{Caesarian section} \\ 0 & \textit{Natural birth} \end{cases}$$

According to Ganesalingam (1989), classification of objects to groups is usually thought of as partition of objects into subsets in which the members are more similar. Classifying individuals into groups such that there is relative homogeneity within the groups and heterogeneity between the groups is a problem which has been considered for many years. For this paper the author intents to re-classify objects into groups in which they were known to belong using logistic regression analysis.

Logistic Model

Let $X_1, X_2, \ldots, X_p$ be predictor variables which consist of qualitative and quantitative variables and Y as a dichotomous dependent variable, where Y is coded as 1 or 0 for its two categories as indicated above.

$$E(X) = \frac{1}{1+\exp[-(B_0 + \sum_{j=1}^{k} B_j x_j)]} \quad \ldots\ldots\ldots\ldots\ldots (1)$$

This equation can be written in a form that describes the probability of occurrence of one of the two possible outcomes of Y, as follow

$$P(Y=1) = \frac{1}{1+\exp[-(B_0 + \sum_{j=1}^{k} B_j x_j)]} \quad \ldots\ldots\ldots\ldots\ldots (2)$$

In general,

$$f(Z) = \frac{1}{1+e^{-z}} \quad \text{where } z = B_0 + \sum_{j=1}^{k} B_j X_j$$

The function f(z) is called logistic function. This function is well suited to modeling a probability, since the values of f(z) varies from $-\infty$ to $+\infty$. The logistic model, therefore, is set up to ensure that, whatever estimate of risk we get, it always falls between 0 and 1. This is not true for other models, which is why the logistic model is often used when a probability must be estimated.

Estimating the Odds Ratio Using Logistic Regression

The regression coefficients B_j in the logistic model given in (1) play an important role in providing information about the relationships of the predictors in the model to the exposure variable. The qualification of this relationship involves a parameter called the odd ratio (Kleinbaum *et.al,* 1998)

The odd ratio is a measure of effect because it is a measure that compares two or more groups in predicting the outcome variable.

The odd of an event D = {Y=1}

$$\text{Odds (D)} = \frac{\Pr(D)}{1-\Pr(D)}$$

Any odds ratio is defined as a ratio of two odds.

The logistic regression is written as

$$\text{Logit}\,[\Pr(Y=1)] = \log_e [\text{odds}(Y=1)]$$

$$= \log_e \left[\frac{\Pr(Y=1)}{1-\Pr(Y=1)}\right].$$

Thus, equation (1) becomes

$$\text{Logit}\,[\Pr(Y=1)] = B_0 + \sum_{j=1}^{k} B_j X_j$$

The odds ratio, say, for groups A and B can be defined as

$$e^{\sum_{j=1}^{k}(X_{Aj}-X_{Bj})B_j}$$

In general the odds ratio for groups A and B in given by

$$\frac{\text{Odds for } X_A}{\text{Odds for } X_B}$$

$$= \frac{e^{(B_0+\sum_{j=1}^{k} B_j X_{Aj})}}{e^{(B_0+\sum B_j X_{Bj})}}$$

$$= e^{\sum_{j=1}^{k}(X_{Aj}-X_{Bj})B_j} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

The constant term B_0 in the logistic model (1) drops out of the odds ratio expression in (3). The expression (3) describes a population odds ratio parameter because the B_j terms in the expression are themselves unknown population parameter. An estimate of this population odds ratio is obtained by fitting the logistic model using maximum likelihood estimation and substitute in the ML estimates B_j, together with the values of X_{Aj} and X_{Bj}, into the formula (3) to obtained a numerical value for the odds ratio.

RESULTS AND DISCUSSION

Data from Health Records of 184 Pregnant women who delivered at the General Hospital, Wuse which comprises of 99 natural birth and 85 caesarian section were used for the study. The Average Age of the Mothers is 27.38yrs with standard deviation 5.4 for Natural birth while the average weight of the Mothers 28.91yrs with standard deviation 4.53 for Caesarian section, the Mothers weight is 69.54kg and 74.11kg respectively, the Mother's height is 161.93m and 160.42m respectively and the Baby's weight is 3.31 for both. The table below depicts the Nagelkerke R square of 8.5% of the total variation in the outcome variable (Natural birth and Caesarian section) explained by the logistic regression model fitted into the data.

Table 1: Model summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	214.942	0.63	0.085

The estimates of the logistic regression model parameters are presented in Table 2 below. The Wald statistics for Mother's height, weight, age, baby's weight and gender are 3.454, 4.179, 1.806, 0.144 and 1.710 respectively. These show that the three are important risk factors to determine the status of mode of delivery for the expectant mothers. The Mother's weight is more important that the age going by the value of the Wald statistic, besides the Wald value for Mother's age, height, baby's weight and gender is not significant while that of mother's weight is highly significant (P-value of 0.041). The odds ratio for Mother's age is 1.043 meaning that an increase in age by one year, Mother's age will increase the rate of the mode of delivery of the expectant mothers by 0.043 (95% C.I. 0.981 to 1.110) .

Table 2: Estimates of the Logistic Regression Model

	B	S.E.	Wald	df	Sig	Exp(B)	95% C.I. for Exp (B) Lower	Upper
Mother's height	-0.039	0.021	3.454	1	0.063	0.962	0.923	1.002
Mother's weight	0.025	0.012	4.179	1	0.126	1.065	1.001	1.050
Mother's age	0.043	0.032	1.806	1	0.179	1.043	0.981	1.110
Baby's weight	-0.110	0.289	0.144	1	0.704	0.896	0.508	1.579
Baby's gender	0.406	0.310	1.710	1	0.191	1.501	0.817	2.757
Constant	3.290	3.570	0.849	1	0.357	26.849		

The overall accuracy of this model (logistic regression) to predicting the mode of delivery of the expectant mothers, with a (predicted probability of 0.5 or greater) is 65.8% as shown in Table 3. The predictive model is z = 3.570 +0.021Mothers height +0.012Mothers weight +0.032Mothers age +0.289Babys Weight + 0.310Babys gender. The interest is to use this model (logistic regression) to predict the outcome for a new case. To determine how good the model is we computed the followings: The sensitivity from Table 3 is 66.18%.

Table 3: Classification Result (for logistic model)

Observed	Predicted 1	0	% Correct
1	76	23	76.8
0	40	45	52.9
Overall %			65.7

65.7% of the original grouped cases correctly classified and specificity is 65.52%. The positive predictive value (PPV) is 65.75% and the negative predictive value (NPV) is 68.70%. These results show that whenever we have a new subject, we can use logistic model to predict her probability of delivering through natural birth. For instance, if given the mothers height, weight, age, baby's weight and gender of an expectant mother and if the predictive model gives a low probability, it means that the subject is very unlikely to deliver through natural birth because the NPV reveals that we should be 68.70% confident and on the other hand if the model gives high probability it means that the subject is very likely to deliver through natural birth because the PPV gives 65.75% confidence.

The summary results obtained under the discriminant analysis are presented in Table 4 and 5 to justify the need to use logistic regression when variables contain both discrete and continuous data.

Table 4: Test for Equality of Group Means (through Discriminant Analysis ANOVA tests)

	Wilks' Lambda	F	Df1	Df2	Sig
Y_1	0.990	1.810	1	182	0.180
Y_2	0.974	4.877	1	182	0.028
Y_3	0.977	4.213	1	182	0.042
Y_4	1.000	0.006	1	182	0.940
Y_5	0.992	1.478	1	182	0.226

The above Table 4 contains the results obtained when discriminant analysis was employed to classify the drug traffickers. In the table the Y_1 represents mothers height, Y_2 the mothers weight, Y_3 the mothers age, Y_4 the baby's weight and Y_5 the baby's gender. All the five variables used in the model except two, Y_2 and Y_3, are significant. This implies that only mothers weight and mothers age can be used to predict whether the mode of delivery of expectant mothers is by natural birth or caesarian section.

Table 5: Classification Results (for Discriminant Analysis)

Observed	Predicted 1	0	% Correct
1	64	35	64.65
0	30	55	64.71
Overall %			64.70

64.7% of original grouped cases correctly classified. Comparing Tables 3 and 5 of classification above the overall percent grouped cases correctly classified is higher when logistic regression model was employ to classify the illicit drug offenders. The positive predictive value is 65.75% while the negative predictive value is 68.70%

Conclusion

The logistic regression model and discriminant analysis were used to classify data collected on the mode of delivery of expectant mothers. The results obtained when logistic regression was employed shows that 65.8% of the original grouped cases correctly classified while discriminant analysis correctly classified 64.7% of the same group. Moreover, the discriminant analysis results indicate that mothers weight and age are significant in classification of the mode of delivery of an expectant mother, but the logistic regression indicate only mothers weight as important variable in the classification.

References

1. Agresti A, 1996. An Introduction to Categorical Data Analysis. John Wiley and Sons. New York.
2. Balogun OS, Akingbade TJ and Akinrefon AA, 2013. Evaluation of Logistic Regression in Classification of drug data in Kwara State. International Journal of Computational Engineering Research, Vol. 3, Issue 3, Page 54-58.
3. Ganesalingam S, 1989. Classification and Mixture Approaches to Clustering via Maximum likelihood. Applied Statistics, 38, no 3, Page. 455 – 466.
4. Hastie T, Tibshirani R and Friedman JH, 2001. The Elements of Statistical Learning. Springer.
5. Kleinbaum DG, Kupper LL, Muller KE and Nizam A, 1998. Applied Analysis and Multivariate Methods. Third Edition. Duxbury Press.
6. UNICEF, (2008): Progress for Children Report
7. Http://www.math.toronto.edu/mathnet)

Cite this article as:
Balogun Oluwafemi Samson, Evaluation of Logistic Regression on Mode of the Delivery of Expectant Mothers, International Journal of Bioassays, 2015, 4 (06), 3990-3993.

Source of support: Nil
Conflict of interest: None Declared