BINARY REGRESSION IN PRACTICE-BASED RESEARCH

Olaniyan Fatai A.
MBBS Ib, MSc Ib (Epid & Med Stat), FWACP, FMCFM
Consultant Family Physician at the University College Hospital, Ibadan, Oyo State, Nigeria

INTRODUCTION

Family medicine is an academic and scientific discipline, with its own educational content, **practice-based research**, evidence-based clinical activities, and a clinical specialty oriented to primary care. Adopting a critical and research-based approach to practice and maintaining this through continuous learning and quality improvement is one of the cardinal features of Family Medicine as a specialty[1]

Family physicians must be familiar with the general principles, methods, concepts of scientific research, and the fundamentals of statistics[1] The generation of new knowledge through practice-based research is a hallmark of good clinical practice in first contact doctors.[1]

In the past, the volume of research in Family Medicine was comparatively low, with the increasing number of Family Physicians and knowledge of the specialty, more articles in the context of our practice are emerging and this continues to influence patient management successes in the subregion.[2,3]

Most of the research relevant to the practice of First Contact clinicians is based on categorical data. The binary variable has two possible outcomes (eg: sex (male/female) or having hypertension (yes/no). A random variable is transformed into a binary variable by defining a “success” and a “failure”. There could also be more than two categories as in R by C contingency tables[4,5,6,]

Logistic regression

Logistic regression is a transformation of the linear regression model that allows us to **probabilistically model binary variables.** It is also known as a generalized linear model that uses a logit link. It is sometimes called the logistic model or logit model analyzes the relationship between multiple independent variables and a categorical dependent variable, then estimates the probability of occurrence of an event by fitting data to a logistic curve. Logistic regression is the most popular multivariable method used in health science. There are two models of logistic regression, binary logistic regression, and multinomial logistic regression. [7,8,9,10]

Binary logistic regression is typically used when the dependent variable is dichotomous and the independent variables are either continuous or categorical. When the dependent variable has more than two categories, a multinomial logistic regression can be employed.[8,9,10,11]

Types of Logistic regression Models

There are three types of logistic regression models, which are defined based on categorical responses.[8,9,10,11]

a. **Binary logistic regression:** In logistic regression models, this is the most commonly used approach and probably most relevant in clinical decision-making in practice-based research. In this approach, the dependent variable is dichotomous in nature—i.e. it has only two possible outcomes, often represented with dummy variables 0 or 1. It is used to predict the membership of only two categories in a model. Examples of its use include predicting success/failure at the Diploma in Family Medicine final Examination; BP reduced/BP not reduced, and transfused/not transfused. [8,9,10,]
b. **Multinomial (polychotomous) logistic regression:** In this type of logistic regression model, the dependent variable has three or more possible outcomes with no implied order. In this type of regression, we want to predict membership of more than two categories (normal BP, Prehypertension BP, and Hypertensive BP[8,9,10]
c. **Ordinal logistic regression:** This type of model is like a multinomial in ranked (implied) order, the response variable has three or more possible outcomes. [8,9,10,11]

Logistic regression is a model for predicting categorical outcomes from categorical and continuous predictors. In practice-based medical research logistic regression is used to generate models from which predictions can be made about the likelihood that a particular event (the levels of risk factors such as high BP, high serum cholesterol, and cigarette smoking.) will bring a particular outcome (risk of developing CHD or not).[11,12]

Based on existing data in the literature, the logistic regression model can be used to establish variables(BP, cholesterol level and cigarette smoking) that can predict (risk of developing CHD or not). These variables can then be measured for new patients and their values placed in the logistic regression model to obtain a

REGRESSION MODEL AND CONCEPTS RELATED TO LOGISTIC REGRESSION

Mathematical equations in regression analysis

$Y = \beta_0 + \beta X + \varepsilon_j$ and εj is the error term

In a simple linear equation when there is one predictor variable gives the equation of the best-fit line in linear regression as shown.

In $\{p / (1-p) = \beta_0 + \beta X + \varepsilon_j$ (εj is the error term)

Taken p as the probability that the event Y occurs, p(Y=1) and p/(1-p) is the "odds ratio"

The logistic distribution constrains the estimated probabilities to lie between 0 and 1.

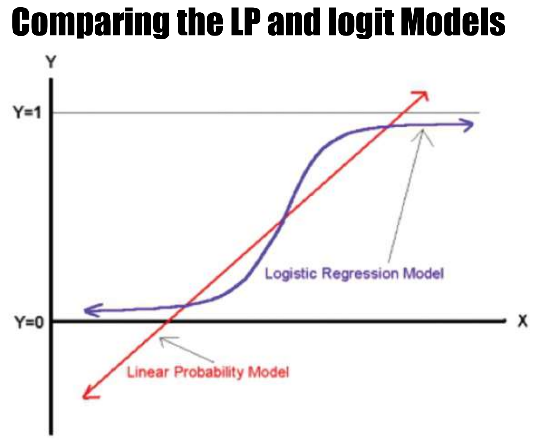
The estimated probability is $p = 1/ [1 + \exp(\beta_0 - \beta_X)]$

if you let $\beta_0 + \beta_X = 0$, then p = .50 as $\beta_{0+}\beta_X$ gets

as $_{0+X}$ gets big, p approaches 1 and as $_{0+\ X}$ gets really small, p approaches 0

An explanation of logistic regression can begin with an explanation of the standard logistic function. The logistic function is a sigmoid function, which takes any real input and outputs a value between zero and one. For the logic, this is interpreted as taking input log odds and having output probability. The standard logistic function is defined as follows:

- The logistic model is interpreted as the probability of the dependent variable equaling success/failure cases/non-cases. It's clear that the response variables are not identically distributed and differ from one data point to another, though they are independent given design matrix and shared parameters as shown below [7,8,11]



Multiple linear regression gives

$Y1 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_{3.} + \beta_j X_j + \varepsilon_j$

when there are two or more.

- β_0 is the value of the outcome when the predictors are zero (the intercept), the bs quantify the relationship between each predictor and outcome, X is the value of each predictor variable and ε is the error in prediction (the residual).

- One of the assumptions of the linear model is that the relationship between the predictors and outcome is linear. When the outcome variable is categorical, this assumption is violated, this was then linearized using the logarithmic transformation. [7, 8, 11]

Logistic regression uses this transformation to express the linear model equation in logarithmic terms (called the logit). In doing so it allows us to predict categorical outcomes using the standard linear model [7, 8, 11]

This logit model analyses the relationship between multiple independent variables and a categorical dependent variable and estimates the probability of occurrence of an event by fitting data to a logistic curve.[7,8,11]

As an illustrative example, consider how coronary hypertension can be predicted by abdominal obesity. The probability of hypertension increases with high BMI, high salt intake, poor exercise, the average number of cigarettes smoked, and abdominal obesity. However, the relationship between hypertension and BMI is nonlinear and the probability of hypertension changes very little at the low or high extremes of BMI. This pattern is typical because probabilities cannot lie outside the range from 0 to 1. The relationship can be described as an 'S'-shaped curve, as shown in the above figure. The logistic model is popular because the logistic function, on which the logistic regression model is based, provides estimates in the range 0 to 1 and an appealing S-shaped description of the combined effect of several risk factors on the risk for an event. [8,11,12]

1. Odds

The odds of an event are the ratio of the probability that an event will occur to the probability that it will not occur. If the probability of an event occurring is p, the probability of the event not occurring is (1-p). Then the corresponding odds is a value given by odds of {Event}= p /(1-p). note that p = pi(π). [4,5,6,12,13]

Since logistic regression calculates the probability of an event occurring over the probability of an event not occurring, the impact of independent variables is usually explained in terms of odds. With logistic regression, the mean of the response variable p in terms of an explanatory variable x is modelled relating p and x through the equation $p={}_0+x$.[7,11,12,13]

Logit form of the model: $\dfrac{p}{1-p} = \dfrac{P(Yes)}{P(No)}$

The logistic model assumes a linear relationship between the predictors and the log (odds).

$$\log\left(\frac{p}{1-p}\right) = b_0 + b_1X$$

logit (y) =ln (odds) =ln (p /1-p) =β_0+ βx, where p is the probability of outcome(dependent) and x is the explanatory(independent) variable. The parameters of the logistic regression are α(or β_0) and β. This is the simple logistic model. [11,12,13]

1. Odd Ratio and Logistic Regression

The odds ratio (OR) is a comparative measure of two odds relative to different events an OR is a measure of association between an exposure and an outcome. The OR represents the odds that an outcome (e.g. disease or disorder) will occur given a particular exposure (e.g. health behaviour, medical history), compared to the odds of the outcome occurring in the absence of that exposure. [6,7,10,11]

OR to compare two groups is to look at the ratio of their odds $odds = \frac{p}{1-p} = e^{b_0+b_1X}$

Odds Ratio =OR is Odds1 / Odds2

If X is replaced by X+1,

then $odds = e^{b_0+b_1X}$ is replaced by $odds = e^{b_0+b_1(X+1)}$

The Odd ratio now becomes

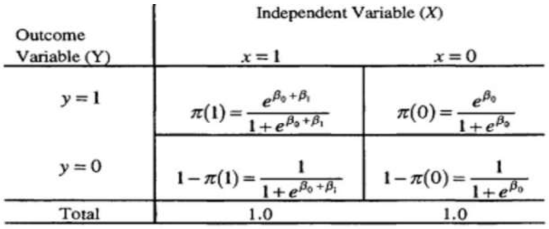
$$\frac{e^{b_0+b_1(X+1)}}{e^{b_0+b_1X}} = e^{b_0+b_1(X+1)-(b_0+b_1X)} = e^{b_1}$$

Then the exponential of beta that is now the odd ratio in logistic regression. [4,8,9,11]

The possible values of the logistic probabilities may be conveniently displayed in a 2 x 2 table as shown below. The odds of the outcome being present among individuals with x = 1 is defined as p(l)/[1-p(l)]. Similarly, the odds of the outcome not being present among individuals with x=O is defined as p(0)/[1-p(0)]. The odds ratio, denoted OR, is defined as the ratio of the odds for x = 1 to the odds for x = 0, and a typical research sample is given by the equation OR= [p(1)/[1-p(1)] / [p(0)/[1- p(0)]. The odds ratio (OR) is the ratio of the odds for x = 1 to the odds for x = 0, and in a population, it is given by the equation below:

$$\mathrm{OR} = \frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]}$$

Substituting the expressions for the LR model OR now gives values of the LR Model with Dichotomous Independent variables. [4,8,9,11]



Outcome Variable (Y)	Independent Variable (X) $x=1$	 $x=0$
$y=1$	$\pi(1)=\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$	$\pi(0)=\frac{e^{\beta_0}}{1+e^{\beta_0}}$
$y=0$	$1-\pi(1)=\frac{1}{1+e^{\beta_0+\beta_1}}$	$1-\pi(0)=\frac{1}{1+e^{\beta_0}}$
Total	1.0	1.0

Calculating the Odd Ratio in a 2 by 2 contingency table given the odd as below:

$$\mathrm{OR} = \frac{\left(\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right)\Big/\left(\frac{1}{1+e^{\beta_0+\beta_1}}\right)}{\left(\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)\Big/\left(\frac{1}{1+e^{\beta_0}}\right)}$$

$$= \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}}$$

$$= e^{(\beta_0+\beta_1)-\beta_0}$$

$$= e^{\beta_1}.$$

The logistic curve

Logistic regression is a method for fitting a regression curve, *y* =f(*x*), when *y* consists of binary coded (0, 1- -failure, success) data. When the response is a binary (dichotomous) variable and *x* is numerical, logistic regression fits a logistic curve to the relationship between *x* and *y*. Logistic curve is an S-shaped or sigmoid curve, often used to model population growth curve. A logistic curve starts with slow, linear growth, followed by exponential growth, which then slows again to a stable rate.

A simple logistic function is defined by the formula. [8,9,11]

$$y = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}}$$

To provide flexibility, the logistic function can be extended to the form $y = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}} = \frac{1}{1+e^{-(\alpha+\beta x)}}$

Logistic regression with a dichotomous independent variable coded **1** and 0 on the Y-axis, the relationship between the odds ratio and the regression coefficient is the exponential of beta.

EVALUATING THE PERFORMANCE OF A MODEL (MODEL FIT)

The probability of the observed results, given the parameter estimates, is used to determine how well the estimated model fits the data.

There are several statistics that can be used for comparing alternative models or evaluating the performance of a single model: [8,11,13]

- **Model Chi-Square**
- **Percent Correct Predictions**
- **Pseudo-R^2**

Model Chi-square

Likelihood index: If the model fits perfectly, the –2LL, will equal 0. Goodness-of-fit statistic (similar to the F test in multiple regression) takes into consideration the difference between the observed probability of an event and the predicted probability—chi-square distribution

In SPSS, rather than reporting the log-likelihood itself, the value is multiplied by –2 (and sometimes referred to as –2*LL*): this multiplication is done because –2*LL* has an approximately chi-square distribution and so makes it possible to compare values against those that we might expect to get by chance alone

Likelihood ratio (LR), the statistic is LR[i] = -2[LL() - LL(,)]

LR[i] = [-2LL (of beginning model)] - [-2LL (of ending model)]}

The LR statistic is distributed chi-square with i degrees of freedom, where i is the number of independent variables. [8,11,15,16]

Use the "Model Chi-Square" statistic to determine if the overall model is statistically significant. The model

	HBP	NBP	Pooled
Ending -2 LL	620.20	360.80	1065.30
Chi-Square	84.30	[Pooled - (HBP + NBP)]	
DF	4		
Critical Value	12.3	p = .02	

If the chi-squared value is greater than the critical value, the set of coefficients is statistically different. The pooled model here is inappropriate[.9,10,14,15]

The Wald statistic: Assessing the contribution of predictors.[8,11,14,15]

The Z statistics is known as Wald statistic, SPSS reports it as z^2 which transforms it so that it has a Chi-square distribution. The z statistics is used to ascertain whether a variable is a significant predictor of outcome, however, one needs to be cautious that regression coefficient (b) is getting large the standard error (SE) tends to be inflated resulting in z- statistic being underestimated. The inflation of the SE increases the probability of rejecting a predictor as being significant when it is contributing significantly to the model (Type II error)

The Wald statistic for the coefficient is: Wald = [/s.e.$_B$]^2, which is distributed chi-square with 1 degree of freedom. [8,11,15,16]

The "Partial R" (in SPSS output) is R = {[(Wald-2)/(-2LL()]}$^{1/2}$

THE PERCENT CORRECT PREDICTIONS

Hosmer Lemeshow test (based on Chi-Square) compares prediction to "perfect model". When not significant, the null hypothesis that the model fits is supported, ie a non-significant result indicates that the model fits, and a significant result indicates that it doesn't.[.9,10,14,15]

By assigning these probabilities 0s and 1s and comparing these to the actual 0s and 1s, the % correct Yes, % correct No, and overall % correct scores are calculated.[11,15,16]

The "Percent Correct Predictions" statistic assumes that if the estimated p is greater than or equal to .5 then the event is expected to occur and not otherwise.

By assigning these probabilities 0s and 1s and comparing these to the actual 0s and 1s, the % correct Yes, % correct No, and overall % correct scores are calculated. [11,15,16]

	Observed	Predicted	Correct (%)
	0	1	
0	300	30	91.00%
1	150	50	33.33%
		Overall	67.89%

PSEUDO-R SQUARE

R^2 values quantify the proportion of the variance explained by the model

One psuedo-R^2 statistic is the McFadden's-R^2 statistic: McFadden's-R^2 = 1 - [LL(,)/LL()]

{= 1 - [-2LL(,)/-2LL()]}

where the R^2 is a scalar measure that varies between 0 and (somewhat close to) 1 much like in the R^2 in a linear model.[11,15,16]

In order to understand how much variation in the dependent variable can be explained by the model (the equivalent of R^2 in multiple regression) and as described in the "Model Summary table": This table contains the Cox & Snell R Square and Nagelkerke R Square values, they are interpreted in the same way, but the most preferred to be reported is the Nagelkerke R^2 value.[11,14,15]

ASSUMPTIONS IN BINARY LOGISTIC REGRESSION

When you choose to analyze your data using binomial logistic regression, part of the process involves checking to make sure that the data you want to analyze has actually met these assumptions. It is only appropriate to use a binomial logistic regression if your data "passes" assumptions that are required for binomial logistic regression to give a valid result.:

1. The dependent variable should be measured on a dichotomous scale. Examples are gender ("males" and "females") and the presence of a disease ("yes" and "no").

2. You have one or more independent variables, which can be either continuous (i.e., an interval or ratio variable) or categorical (i.e., an ordinal or nominal variable).

3. You should have independence of observations and the dependent variable should have mutually exclusive and exhaustive categories.

4. There needs to be a linear relationship between any continuous independent variables and the logit transformation of the dependent variable.[6,7, 11,15,16]

It is better to check the assumptions in that order. If a violation to the assumption is not correctable, then it is invalid to use a binomial logistic regression and if you do not run the statistical tests on these assumptions correctly, your result may not be valid.[11, 15,16]

Procedure for Analysis in SPSS Statistics [6,15,16]

To analyse your data using a binomial logistic regression in SPSS Statistics when none of the assumptions have been violated. At the end of these steps, we try to interpret the results from the binary logistic regression. Doing the analysis on SPSS runs thus:

Click Analyse > Regression > Binary Logistic... on the main menu. This will take you the Logistic Regression dialogue box.

Then transfer the dependent variable into the Dependent: box, and the independent variables into the Covariates: box, using the arrow buttons

As for standard logistic regression, you should ignore the previous and next buttons because they are for sequential (hierarchical) logistic regression. Keep the method option at the default value, which is the "enter"

Click on the categorical button, this will take you to Logistic regression and defined categorical variables.

Then define all the categorical predictor values in the logistic regression model. It does not do this automatically. Transfer the categorical independent variable from the covariates on the left to the categorical covariate box on the right side.

In the change contrast area, you may want to change the reference category from the last option to the first option or the other way round then, click on the change button. If you are to compare males to females the females are often used as the reference category and coded "0". You then click on the continue button, which takes you back to the Logistic Regression dialogue box. Clicking on the options button will present you with the Logistic Regression: Options dialogue box

In the Statistics and Plots area: select classification plot, Hosmer Lemeshow goodness-of-fit, Casewise listing of residuals, CI for exp(B), the outliers outside 2 standard deviations, and the display at the last step. Also include constant in the model.

When you click the continue button it takes you back to the Logistic Regression dialogue box. Then Click on the OK button, the output involves the processing of graphics, so this takes a little bit of time to generate the final output.

Method selection allows you to specify how independent variables are entered into the regression analysis. Using different methods, you can construct a variety of regression models from the same set of variables.

The enter (Regression) or Forced entry selection. A procedure for variable selection in which all variables in a block are entered in a single step.

Stepwise (Regression). At each step, the independent variable not in the equation that has the smallest probability of F is entered, if that probability is sufficiently small. Variables already in the regression equation are removed if their probability becomes sufficiently large. The method terminates when no more variables are eligible for inclusion or removal. [6, 11,15,16]

INTERPRETATION AND REPORTING LOGISTIC REGRESSION

The interpretation of the odds ratio depends on whether the predictor is categorical or continuous. An OR of 1 means that nothing is going on or no difference. Odds ratios that are > 1 indicate that the event is more likely to occur as the predictor increases. Odds ratios that are < 1 indicate that the event is less likely to occur as the predictor increases.

When the prevalence of a disorder is low (under about 10%), the RR and OR are nearly the same.

As the prevalence increases, the OR becomes larger than the RR. In fact, the OR sets an upper bound for the RR. [8,11,16]

ANALYSIS OF RESIDUAL

On the SPSS: From the Data View, there are three new variables that have been created. The first is the predicted probability of that observation and is given the variable name of PRE_1. The second variable contains the raw residuals (the difference between the observed and predicted probabilities of the model) and is given the variable name of RES_1. The third variable has standardized residuals based on the raw residuals in the second variable and will be given the variable name as ZRE. [1,11,12,13,16]

Click Graphs; Drag the cursor over the Legacy Dialogs drop-down menu; Click Scatter/Dot., Click Simple Scatter to select it; Click Define, then on the RES_1 or raw residual variable to highlight it.

Click on the arrow to move the variable into the Y Axis: box, then on the PRE_1 or predicted probability variable to highlight it.

Click on the arrow to move the variable into the X-Axis: box, then click OK. [11, 15, 16]

THE STEPS FOR INTERPRETING THE SPSS OUTPUT FOR A LOGISTIC REGRESSION

Scroll down to Block 1: Method = Enter section of the output. Look in the Omnibus Tests of Model Coefficients table, under the Sig. column, in the Model row. This is the p-value that is interpreted.

If the p-value is < .05, then researchers have a significant model that should be further interpreted but If the p-value is > .05, then researchers do not have a significant model and the results should be reported.

Look in the Hosmer and Lemeshow Test table, under the Sig. column. This is the p-value you will interpret.

If the p-value is < .05, then the model does not fit the data, if the p-value is >.05, then the model does fit the data and should be further interpreted.

Look in the Classification Table, under the Percentage Correct in the Overall Percentage row. This is the total accuracy of the model. Researchers want it to ultimately be at least 80%.

Look in the Variables in the Equation table, under the Sig., Exp(B), and Lower and Upper columns. The Sig. column is the p-value associated with the adjusted odds ratios and 95% CIs for each predictor variable. The value in the Exp(B) is the adjusted odds ratio. The Lower and Upper values are the limits of the 95% CI associated with the adjusted odds ratio.

Researchers will interpret the adjusted odds ratio in the Exp(B) column and the confidence interval in the Lower and Upper columns for each variable. If the confidence interval associated with the adjusted ratio crosses over 1.0, then there is a non-significant association. The p-value associated with these variables will also be > .05.

If the adjusted odds ratio is > 1.0 and the CI is entirely above 1.0, then exposure to the predictor increases the odds of the outcome. If it is < 1.0 and the CI is entirely below 1.0, then exposure to the predictor decreases the odds of the outcome.

We also need to note that if the variable is measured at the ordinal or continuous level, then the adjusted odds ratio is interpreted as "for every unit increase" in the ordinal or continuous variable, the risk of the outcome increases at the rate specified in the odds ratio.[11, 15, 16]

Residuals and logistic regression

Residuals are the error associated with predicting or estimating outcomes using predictor variables. Residual analysis is extremely important for meeting the linearity, normality, and homogeneity of variance assumptions of logistic regression, so we need to conduct residual analysis . There are many types of residuals such as ordinary residual, Pearson residual, and studentized residual. They all reflect the differences between fitted and observed values and are the basis of varieties of diagnostic methods. [1] A basic type of graph is to plot residuals against predictors or fitted values. If a model is properly fitted, there should be no correlation between residuals and predictors and fitted values.[12,13]

REPORTING GUIDELINES

There are no generally accepted ways of reporting the results of logistic regression. Some use a table similar to the one for ordinary least squares regression, except for replacing the t-test with the Wald statistic and its df and adding an extra column called either Odds Ratio (OR) or Exp(b) and the p-value. In family medicine research we commonly add the confidence interval around Exp(b). Whichever format you use, be sure to report one or more of the pseudo-R^2 statistics for logistic regression; The two mostly recommended are McFadden's adjusted R^2 and Nagelkerke's R^2. Others like Cox and Snell's statistics etc. are hardly seen in literatures. [6,711,16]

CONCLUSION

Binary Logistic regression is a type of multivariable analysis used with increasing frequency in Family Medicine practice-based research because of its ability to model the relationship between a dichotomous dependent variable and one or more independent variables.

REFERENCES

1. EURACT. The European definition of general practice/family medicine. Barcelona: WONCA Europe. 2002. woncaeurope.org/file/3 b13bee8-5891-455e-a4cb a670d7bfdca2/Definition%20EURACTshort%20version%20revised%202011.pdf
2. WONCA E. The European definition of general practice/family medicine. http://www.wonEurope.org/.2005.
3. Hummers-Pradier E, Beyer M, Chevallier P, Eilat-Tsanani S, Lionis C, Peremans L, Petek D, Rurik I, Soler JK, Stoffers HE, Topsever P. The Research Agenda for General Practice/Family Medicine and Primary Health Care in Europe. Part 1. Background and methodology1. The European journal of general practice. 2009 Jan 1;15(4):243-50.
4. Liang KY, Zeger SL, Qaqish B. Multivariate regression analyses for categorical data. Journal of the Royal Statistical Society: Series B (Methodological). 1992 Sep;54(1):3-24.
5. Preisser JS, Koch GG. Categorical data analysis in public health. Annual Review of Public Health. 1997;18(1):51-82.
6. Norman GR, Streiner DL. Biostatistics: the bare essentials. BC Decker. Inc., Hamilton, Ontario. 2000.
7. Kirkwood BR, Sterne JA. Essential medical statistics. John Wiley & Sons; 2010 Sep 16.
8. Lemeshow S, Sturdivant RX, Hosmer Jr DW. Applied logistic regression. John Wiley & Sons; 2013.
9. AJ, Patterson CC, Raines B. On the use of a logistic risk score in predicting risk of coronary heart disease. Statistics in medicine. 1990;9(4):385-96.
10. Peng CJ, Lee KL, Ingersoll GM. An introduction to logistic regression analysis and reporting. J Educ Res. 2002. 96(1):3–14.
11. Park HA. An introduction to logistic regression: from basic concepts to interpretation with particular attention to nursing domain. Journal of Korean Academy of Nursing. 2013;43(2):154-64.
12. Alsharif AA, Pradhan B. Urban sprawl analysis of Tripoli Metropolitan city (Libya) using remote sensing data and multivariate logistic regression model. Journal of the Indian Society of Remote Sensing. 2014;42(1):149
13. LaValley MP. Logistic regression. Circulation. 2008 May 6;117(18):2395-9.
14. Menard S. Applied logistic regression analysis. 2nd ed. New York: SAGE Publications; 2001:1
15. Meyers LS, Gamst GC, Guarino AJ. Performing data analysis using IBM SPSS. John Wiley & Sons; 2013.
16. Field A. Discovering Statistics using IBM SPSS Statistics. SAGE; 2013.