



## Logit model estimates of CSR and performance measurement companies in Indonesia

Ahmad Subagyo<sup>1</sup>, Teguh Sugiarto<sup>2,\*</sup>

<sup>1</sup>Lecturer Management of STIE GICI, Depok, Indonesia

<sup>2</sup>Lecturer at Budi Luhur and AAJ, Jayabaya University, Jakarta, Indonesia

### ARTICLE INFO

#### Article history:

Received 9 December 2015

Received in revised form

24 February 2016

Accepted 26 February 2016

#### Keywords:

Logit model

CSR

Financial performance

Panel data

### ABSTRACT

This study aims to show how to use different commands to the data analysis company with CSR performance ratio as a variable, in some types of mining companies in Indonesia. To postulate a meaningful modeling, the authors review the literature of the past which can provide references from this research model. From the results it is shown that the probability is not significant and the value of adjusted R squared very low, but that does not mean a regression that was made to produce a spurious regression. For the final result can be explained that the actual level of risk with what is expected, the low and high level. The author uses some of the opinions of previous research in making a conclusion which describes the categorical regression models election, parameter estimation and goodness-of-fit statistic is not enough. The author hopes that further research could give the feel and the better models and logical in using various commands in the data analysis.

© 2016 IASE Publisher. All rights reserved.

## 1. Introduction

In a study done by Bagley et al (2001) and Katz, (1999) regarding the logistical regression, they argued that researchers need to look and give their attention to something bigger, particularly regarding guidance on the use and application of LR model in the results of their research. The existence of several multivariate models and methods of statistical analysis which is considered often appears in several different Sciences, makes an excellent nuance in the application of the analysis of the logistics. The existence of the use of the name "multivariate analysis" or "multivariate analysis" can be used as material for the turn in some existing literature. In fact if we see further, multivariate analysis gives a more specific sense to refer to the existence of togetherness in predicting the results of the analysis of multiple variables as well as multivariate analysis using several variables in predicting the results of that nature. Of several concepts exist, can be set as the basis for a transformation of the logit model and opportunities as well as the existence of the use of the logistic curve or model assumptions that fit. In terms of reporting and interpreting can be set as material in the warn that would be spelled out in the research results. The existence of a number of studies that use the LR in reporting the results of

their studies, can be set as the sample size where a bit small, as well as the need to question the accuracy of the regression models that have been built. From the results of their studies, providing validation that this is considered good and were able to use.

There are several methods and models in the multivariate analysis, the first model that consists of two purposes: the first use of functions of this model is able to make an estimate which is the dependent variable consists of explaining the new dent in the top of the value of the independent variable in the produce, and the second one of this model is able to do in explaining what exactly the relative contribution of each independent variable toward dependent variable upper function, especially with regards to the processes for controlling other independent variables. In making a model, then it should at be sure that inside there are models that describe the relationships and disclose the value of a variable that will do predict. As an example there is a variable as a manifestation of real results from some of the number of products that have been produced or any product that has been produced the calculation of the value of the coefficient and its nature serves as the independent variable. The next, there was a number of coefficients are obtained as the basis for the calculation of process similarities of mathematics, for a specific model. Although the figures will coefficient of yesteryear shows how the impact of each independent variable from within the model in the proposal, from each variable results

\* Corresponding Author.

Email Addresses: [bgv2000@yahoo.com](mailto:bgv2000@yahoo.com) (A. Subagyo),  
[teguh\\_cpconsulting@yahoo.com](mailto:teguh_cpconsulting@yahoo.com) (T. Sugiarto)

have been doing adjustments on each independent variable. If we make a conclusion, then the multivariate method is considered as a tool that can explore a relationship between clarity of two or more variables that are examined, which are as independent variables and the results of prediction will depend from other variables. There are actually four multivariate analysis methods that can be used in analysis or statistical data, including a simple linear regression, logistic regression, discriminant analysis and the final regression analysis which is proportional.

According to Tetrault et al (2008) explaining how the usability as well as the intent of the multivariate analysis model 4 above. The first thing to note is they have a lot of similarities in terms of mathematical explanation, but is considered to have a difference in problems expressing and do a format of the results of a regression in the proposal. In linear regression, there is an explanation that describes how the result variable has the quantity continues. The logistic regression model, then the variable result will usually described with the form or in the form of binary events, such as something opposite to or in pairs but have something function at the beginning and the end. Model analysis of discriminant analysis, the next a results variable analysis will be described with a model analysis of the category or group that its nature is the subject of private owned. To study the use of model analysis of logistic regression (LR) will be presented and displayed by using the basic concepts for the interpretation. But not a possibility of the existence of a benchmarking process for the use of the actual function of the logistic regression, so what is the criteria that is specified can be funded and used in reporting later.

## 2. Theory estimation method

In a study performed by Spicer (2004) explains how we are to understand how exactly the application from a binary logistic regression. Binary logistic regression is one method of analysis that can be used, if we have some variable which is the dependent variable is binary or dichotomous. With a sense that we will be able to see the road exploring, as well as see how influence over the application, which is relatively sustainable because the variables that use a categorical independent variable-shaped. While the dependent variables that we have would function as a tool of analysis of the effect of the interaction that occurs between the independent variables and the dependent.

Suppose we'll do an analysis, with the road makes a hypothetical as described above, then we will do some test of OLS regression which can be applied to the financial data we have, with the dependent variable is a binary variable which is shaped. Regression analysis in logistics has a binary variable, later will produce some estimate probability results with these variables with categorize step labeled 0 and 1. The use of numbers in binary logistic regression analysis such as in

question will be able to bring up the probability that a value represented by the opportunities that will be featured. There is a possibility that will appear with a probability value over a category that has been given, making the existence of the process of designation numbers 1 to 4. In other words, if there are 4 times more likely to arise in the application of the regression analysis were made, in addition there is a designation number 1 than in category 0. The occurrence of the transformation process in the binary regression, can make the problem being solved, and need to be in the know that the existence of the OLS regression in form will be exposed to this form of data which have a binary nature or variables that will appear like the equation likely (odds + 1). The purpose of the analysis is the possibility of being transformed into a number of nature-shaped log, log, as well as the possibility of the logits of nature, or in mean something to force thousands of data that has been raised to generate opportunities. If we describe in the form of the equation will appear, as follows:

$$\text{Predicted logit of } Y = [\beta_1 X_1 + \beta_2 X_2 + \dots + \varepsilon]$$

By looking at the model's equation above, then you can be sure that the logistics of the coefficients, can be determined from the existence of the process optimization of predictive coefficients in and maximum. The intent of the optimization or maximize the value of the coefficients are predictive, will be able to make a force over predictive coefficients, while in regard to maximise the value of can be seen from the application of some of the criteria that are considered to be achieving results objectives are reached. If we look at the specifics, then the predictive process and maximum of yesteryear will result in regression analysis model in the form of logistics or in other words the existence of a likelihood of probability values in hope will happen, so the level of accuracy in the placement of the study as well as the categories in minimizing the difference between the several studies performed. Studies that use elements of the estimates as well as the use of the logarithmic function in the method of analysis, making materials for the collection of value or numeric probability in produce or estimated into the category actually became more effective and efficient. One way in maximizing the value of the coefficient, then the unambiguous or addition coefficient logistics.

In Page et al (2003) describes how the use of independent variable all together and continuously, which eventually will be properly distributed, then the discriminant analysis model can be used in this analysis or widely in used this. In their writings that mention that there are terms and use of the dependent variable name which is more quantitative, or a normal call with analysis of one-Way analysis of variance. This analysis is considered proper can be used in some models of analysis, including in the meet one of the requirements in the data analysis. Some of the methods concerned, where one of the variables used in the analysis, there

is the dependent variable term which is dichotomous or categorical, then some of the things that must be done in the data analysis and related methods is to create some variables. Their intentions to make the variable mixture that is to explain where the variables are continuous and have thought there was a categorical nature. Logistic regression analysis in multiple allowed when an event occurs or where the independent variable always appears but is not distributed with good data. In Spicer (2004) and Garson (2009), it is explained how measurements of some of the samples that are deemed necessary and considered reliable, so the use of logistic regression analysis or the binary method is reliable or necessary, compared with the usual methods of analysis or regression (OLS). They suggest should be able to provide input towards the study discussed based upon the recommendation of a minimum of 50 cases per independent variable, and according to them the opposite must also refer to the recommendations of the 10 cases per independent variable.

Here a few things considered Peng et al (2002) and Spicer (2004) should be able to or should we fill and can be used as material recommendations in data analysis particularly in the logistic regression analysis. The first one namely overall evaluation analysis against a model that in the proposal. The intent here is to do the steps the evaluation analysis indicated with the use of a logistic analysis followed by what is on the diusul model, its independent variable does not affect the dependent variable as a whole. Or with the conclusion of another independent variable used in the analysis models considered can increase the ability of the results of the analysis that we have, in obtaining the value of the dependent variable of the prediction. Thing or this method we can do by way of comparing and not by way of guessing that all of the expected results on the study performed will result in or fall into the category of the most common results.

This analysis tool we can use as there use the ratio likelihood test and other test scores, such as the analysis of the p-value is used by looking at the numbers p value smaller than 0.05 analysis tools are used, these models already demonstrate that the independent variable will be able to make influence the dependent variables against it or not. The second, namely the use of statistical analysis tools for the prediction of perindividu variables in doing research or analysis such as t-test, chi-squared, wald test, which uses a prediction tool for benchmarking the p-value is less than the level of alpha used 0.05 and considered the result is significant. The third is the existence of the use of test data to statistical goodness of fit. The purpose of this test is to look at a model of kindness done the analysis, by way of looking at the accuracy of the models in the proposal, what is the model that made fit, whether the model that made perfect, whether model made no lancung and others. The use of this test is usually assessed with tests Hoesmer-Lemeshow, namely by means of Chi2 number or see the value of which is considered

an indicator of the goodness of the model in the proposal. Usually the same as the previous test, a benchmark that is used is the value or number p then 0.05 was considered > data or models that performed the analysis showed good data. The last use of probability analysis namely yng in estimate. This last analysis model, usually done to see the results of the accuracy of prediction models or equations that we suggested in a study. The model of this analysis will be presented in the form of a table that has been classified. Where the results of a prediction model that has been on the proposal will be compared with the value of the actual test results. Some analysis tools that are used to test this model is that according to Garson (2009), while the Hosmer-Lemeshow test suggested above this model's goodness with Chi2 test as well as for the good model with test-fit. In General for this study only using analysis tools that are present between the existing 3 1 described below.

Their specifications in use binary models will do the premises typing the name of the binary dependent variable, but it must be followed by a list of regressors, in other words we can enter the name into the equation that will be used. Below are three models of binary equation or the writer prefers to call categorical equation is much in use in the distribution of the data that is categorical :

Probit:

$$P_r(y_i = 1|x_i, \beta) = 1 - \Phi(-x_i'\beta) = \Phi(x_i'\beta)$$

$\Phi$  is the cumulative distribution function of the standard normal distribution.

Logit:

$$P_r(y_i = 1|x_i, \beta) = 1 - \left( \frac{e^{-x_i'\beta}}{(1 + e^{-x_i'\beta})} \right) \\ = \frac{e^{x_i'\beta}}{(1 + e^{x_i'\beta})}$$

which is based upon the cumulative distribution function for the logistic distribution.

Extreme value (Gompit):

$$P_r(y_i = 1|x_i, \beta) = 1 - \left( 1 - \exp(-e^{-x_i'\beta}) \right) \\ = \exp(-e^{-x_i'\beta})$$

which is based upon the CDF for the Type-I extreme value distribution. Note that this distribution is skewed.

### 3. Research method

#### 3.1. Data and time research

The author conducted this research using the data with the following criteria: a) With the kind Mining Company listed on the Indonesia Stock Exchange until the date of publication of the annual report as of December 31, 2014. b) Mining

companies always publish annual financial Belief until December 31 2014. c) Mining Company owns the data / information about their CSR disclosure in the financial statements had been audited. To undertake this work on time in February 2016.

### 3.2. Logit methodology Analysis

Binary logistic regression estimates the probability that a characteristic is present, given the values of explanatory variables, in this case a single categorical variable ;  $\pi = Pr(Y = 1|X = x)$ . Probability of success will depend on levels of the risk factor. But this research variabel Y not have caracteristic and not single categorical. Such as : *Variables*: Let  $Y$  be not a binary response variable and  $X = (X_1, X_2, ..., X_k)$  be a set of explanatory variables which can be discrete, continuous, or a combination.  $x_i$  is the

observed value of the explanatory variables for observation  $i$ . In this section of the notes, we focus on a single variable  $X$ . That explanatory variable  $X$  is Financial ratio or spesific performance ratio.

### 4. Result and discussion

In the analysis model created authors use variables and variable financial ratios company CSR in mining companies in Indonesia, displays the information in the status bar. For this type of software that is used by the author, will be given the independent variables of 0 and a numbering scale, as a starting point for an examination of the results of the analysis will be. Here are the results of the analysis in the can contained on the author in Table 1.

**Table 1:** Result Binary Logit Method

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.440124	1.857186	-0.236984	0.8127
CASH_RATIO	-0.997449	0.676734	-1.473917	0.1405
CSRDI	3.273274	3.046085	1.074584	0.2826
LN_TOTAL_ASET	0.022902	0.094160	0.243222	0.8078
TRANSFORMASI_REVERSE_SCO	1.28E-10	9.62E-11	1.326426	0.1847
Dependent Variable: MEDIA_EXPOSURE				
Method: ML - Binary Logit (Quadratic hill climbing)				

From the information, Table 1 above shows the basic information on the estimation technique which is derived from the value of Maximum likelihood (ML) for maximum likelihood and sample the variables used in the equation that is in use estimations, the above information in Table 1, also gives an account of the number of iterations required for a convergence value, and the method used to calculate the coefficient covariance matrix. In Table 1, there are a number of coefficient estimates,

standard errors asymptotically, z-statistic and corresponding p-values. Interpretation of the results of a complex coefficient values sediit and memunyai a fact that the estimated coefficients of binary models can not be interpreted as a marginal effect on the dependent variable. In addition to the summary statistics of the dependent variable, the statistical software Also presents the following summary statistics (Table 2).

**Table 2 :** Result R squared Binary Logit Method

McFadden R-squared	0.090086	Mean dependent var	0.725000
S.D. dependent var	0.452203	S.E. of regression	0.454600
Akaike info criterion	1.320366	Sum squared resid	7.233130
Schwarz criterion	1.531476	Log likelihood	-21.40732
Hannan-Quinn criter.	1.396696	Deviance	42.81463
Restr. deviance	47.05350	Restr. log likelihood	-23.52675
LR statistic	4.238872	Avg. log likelihood	-0.535183
Prob(LR statistic)	0.374642		
Obs with Dep=0	11	Total obs	40
Obs with Dep=1	29		

On the results (Table 2) looks there are some test results descriptive statistics which is not foreign to us look like: mean and standard deviation which is the result of the analysis of the dependent variable, and then there are also the standard error of regression, and residual sum of squares. The latter two measures are calculated in the usual manner using ordinary residuals:

$$e_i = y_i - E(y_i|x_i, \beta) = y_i - (1 - F(-x_i'\beta))$$

There are several other explanations regarding the statistical test results are likely a bit basic, namely: • Log possibilities is the maximum value of the log likelihood function • Average. log • restr. • The LR log statistics • Probability (LR stat) • McFadden R-squared • "Criteria for Information". For further discussion.

In detecting Dependent Variable Frequency analysis models will display the frequency and cumulative frequency table for binary dependent variable in the model. Their naming by the term

categorical regressor statistics, displaying descriptive statistics with the average value and standard deviation for each regressor. Here we see the results in Table 3 the value of descriptive

statistics calculated for the entire sample, and the sample is broken down by the value of the dependent variable on each variable:

**Table 3:** Result Categorical Descriptive Statistic for Explanatory Binary Logit

Variable	Mean		
	Dep=0	Dep=1	All
C	1.000000	1.000000	1.000000
CASH_RATIO	0.717281	0.556843	0.600964
CSRDI	0.302020	0.365517	0.348056
LN_TOTAL_ASET	20.56700	21.38618	21.16091
TRANSFORMASI_REVER SE_SCO	2.94E+09	4.36E+09	3.97E+09
Variable	Standard Deviation		
	Dep=0	Dep=1	All
C	0.000000	0.000000	0.000000
CASH_RATIO	0.791346	0.903264	0.866948
CSRDI	0.082375	0.172500	0.154686
LN_TOTAL_ASET	2.800055	5.158992	4.610412
TRANSFORMASI_REVER SE_SCO	2.05E+08	8.69E+09	7.39E+09
Observations	11	29	40

Seen in Table 3, Predicted results view displays the results of the classification is right and wrong based on user-specific prediction rules, as well as the calculation of the expected value. Results in Table 3 made that observation on the data to be classified as having a probability prediction located above or below the specified cutoff value. With input cutoff value generated four tables in the window equation restrictions. Where the result that the intent by the

author can be seen in Table 4 in which the contingency table of the predicted response is classified on the dependent variable are observed. Here are the results in Table 4 for two tables and related statistics illustrate the results of classification based on cutoff grades:

**Table 4:** Result Expectation-Prediction Evaluation Binary Logit

Expectation-Prediction Evaluation for Binary Specification						
Success cutoff: C = 0.5						
	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)≤C	2	1	3	0	0	0
P(Dep=1)>C	9	28	37	11	29	40
Total	11	29	40	11	29	40
Correct	2	28	30	0	29	29
% Correct	18.18	96.55	75.00	0.00	100.00	72.50
% Incorrect	81.82	3.45	25.00	100.00	0.00	27.50
Total Gain*	18.18	-3.45	2.50			
Percent Gain**	18.18	NA	9.09			

In Table 4 The above described observations concerning the classification for the numbers as the numbers predicted probabilities, which are in the cutoff upper limit or lower limit specified value. In this case the default value has been set at 0.5. The process of classification, observations using a sample proportion of observations. Can believe that the probability value is calculated from the estimated model that only includes an intercept term, C. Figures classification is obtained when the probability is estimated to be less than or equal to the cutoff and observable, or when the predicted probabilities greater than the cutoff and observed. For the above case, 40 of Dep = 0 observations and

28 of Dep = 1 observations correctly classified by the model estimation.

For needs to know that in the literature of statistics, which tell the tables expectation-prediction is referred to as the classification table, but overall, the estimation model correctly predicts 18:18% and 96.55% of observations, that is where the 18:18 of Dep = 0 and 96.55 of Dep = 1 observations. This model is expected to increase at Dep = 1 forecast by 75 percent, but no worse than Dep = 0 prediction (18:18 per cent). Overall, the estimated equation is 18:18 percent better in predicting the response of constant probability model. This change represents an increase of 25

percent compared to 75 percent correct prediction of the standard model. Here we look at Table 5 for

window contains the results of prediction equations based on the analog expected value calculation:

**Table 5:** Result Binary Logit Estimated Equation and Constan Probability

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	3.82	7.18	11.00	3.03	7.98	11.00
E(# of Dep=1)	7.18	21.82	29.00	7.98	21.03	29.00
Total	11.00	29.00	40.00	11.00	29.00	40.00
Correct	3.82	21.82	25.64	3.03	21.03	24.05
% Correct	34.74	75.24	64.10	27.50	72.50	60.12
% Incorrect	65.26	24.76	35.90	72.50	27.50	39.88
Total Gain*	7.24	2.74	3.98			
Percent Gain**	9.98	9.98	9.98			
*Change in "% Correct" from default (constant probability) specification						

In the left-hand of Tabel 5, we compute the expected number of  $y=0$  and  $y=1$  observations in the sample. For example,  $E(\# \text{ of Dep}=0)$  is computed as:

$$\sum_i P_r(y_i = 0 | x_i, \beta) = \sum_i F(-x_i' \beta)$$

where the cumulative distribution function to normal, logistics, or the extreme value distribution.

In the bottom right table, we calculate the expected number of observations for the estimation model and the only constant. For this limited model,  $E(\# \text{ of Dep} = 0)$  is calculated as, where the sample proportion of observations. Between 11 and 29 companies with, the expected number of observations in the estimation model are 34 and 75.

Among the 11 observations, the expected number of observations is 9.98. Can we say that these figures represent approximately 34.74 percent (75.24 percent) improvement over constant probability model. The existence of an Goodness-of-Fit test allows us to perform tests type Pearson goodness-of-fit. There are two goodness-of-fit test: Hosmer (1989) and Andrews (1988a, 1988b). The underlying idea of this test is to compare the expected value that is placed on the actual value by the group. If the difference is "great", we reject the model as provide sufficient fit to the data.

By default the Hosmer-Lemeshow test will conduct a process of grouping, the default grouping method is to form decile. The test results using a standard specification given by Table 6.

**Table 6:** Result Goodnes of fit Binary Specification

Goodness-of-Fit Evaluation for Binary Specification Andrews and Hosmer-Lemeshow Tests								
	Quantile of Risk		Actual	Expect	Actual	Expect	Obs	H-L Value
	Low	High						
1	0.2462	0.5097	2	2.37505	2	1.62495	4	0.14579
2	0.5249	0.6464	2	1.66058	2	2.33942	4	0.11862
3	0.6546	0.6734	0	1.33303	4	2.66697	4	1.99932
4	0.6746	0.7178	1	1.18644	3	2.81356	4	0.04165
5	0.7231	0.7339	2	1.08435	2	2.91565	4	1.06075
6	0.7360	0.7749	1	0.95145	3	3.04855	4	0.00325
7	0.7784	0.8135	2	0.81218	2	3.18782	4	2.17980
8	0.8214	0.8354	1	0.67908	3	3.32092	4	0.18268
9	0.8387	0.8600	0	0.58684	4	3.41316	4	0.68774
10	0.8603	0.9661	0	0.33099	4	3.66901	4	0.36085
		Total	11	11.0000	29	29.0000	40	6.78045
H-L Statistic			6.7805		Prob. Chi-Sq(8)		0.5605	
Andrews Statistic			15.9007		Prob. Chi-Sq(10)		0.1025	

In Table 6, at the top there is a column called "quantiles Risk", in which the column describes the high and low values predicted probabilities for each decile. As also in that column describes the number of actual and expected observations in each group, as well as the contribution of each group to the overall result Hosmer-Lemeshow (H-L) statistical value-great indicates a big difference between the actual value and expected to decile it. The function of a residual public can easily obtain a vector score by

multiplying the residual public by each regressors in. This test can be done with LM specifications, see Chesher et al (1985), and Gourieroux et al (1987).

## 5. Conclusion

According to Long (1997) an appropriate regression model is determined primarily by the level of measurement categorical dependent variable of interest. While in Long (1997), it is described the

parameter estimates and goodness-of-fit statistic is not enough. Therefore, by using two grounds of researchers above, the research that has been done by this author can be concluded that the logit analysis model for measuring financial ratios and CSR in companies in Indonesia is right. From the results displayed in Table 1 to Table 6 shows that the value of probabilistic in significant Table 1 and the value of adjusted R squared very low, but that does not mean a regression that was made to produce a spurious regression. In Table 3 categorical value for the number 0 in the can 11 and to number 1 by 29, but the level of accuracy at yield of 18% for the number 0 and 96% for the number 1 on Tables 4 and 5. For the final result can be explained that the actual level of risk with what is expected, the low and high level (Table 6).

## References

- Andrews DWK (1988a). Chi-Square diagnostic tests for econometric models: theory. *Econometrica*, 56(6): 1419–1453.
- Andrews DWK (1988b). Chi-Square diagnostic tests for econometric models: introduction and applications. *Journal of Econometrics*, 37(1): 135–156.
- Bagley SC, White H and Golomb BA (2001). Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. *Journal of Clinical Epidemiology*, 54(10): 979-985.
- Chesher A, Lancaster T and Irish M (1985). On detecting the failure of distributional assumptions. *Annales de L'Insee*, 59/60: 7–44.
- Garson GD (2009). Logistic regression from *statnotes: topics in multivariate analysis*. Retrieved 6/5/2009 from <http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm>.
- Gourieroux C, Monfort A, Renault E and Trognon A (1987). Generalized residuals. *Journal of Econometrics*, 34(1-2): 5–32.
- Hosmer DW and Stanley Lemeshow (1989). *Applied logistic regression*. 1<sup>st</sup> Edition, John Wiley & Sons, New York.
- Katz MH (1999). *Multivariable analysis: a practical guide for clinicians*. 1<sup>st</sup> Edition, Cambridge University Press, USA.
- Long JS (1997). *Regression models for categorical and limited dependent variables*. Advanced Quantitative Techniques in the Social Sciences Number 7. Sage Publications: Thousand Oaks, CA.
- Page MC, Braver SL and MacKinnon DP (2003). *Levine's guide to SPSS for analysis of variance*. 2<sup>nd</sup> Edition, Taylor & Francis, New Jersey.
- Peng CYJ, Lee KL and Ingersoll GM (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96(1): 3-14.
- Spicer J (2004). *Making sense of multivariate data analysis*. Sage Publications, California.
- Tetrault JM, Sauler M, Wells CK and Concato J (2008). Reporting of multivariable methods in the medical literature. *Journal of Investigative Medicine*, 56(7): 954-957.