Penerbit Akademia Baru

Journal of Advanced Research Design
ISSN (online): 2462-1943 | Vol. 21, No. 1. Pages 6-9, 2016

Formulation of Logit Regression

A. Asraf, S. Sorooshian and J. K. Cheng*

Faculty of Industrial Management, Universiti Malaysia Pahang, 26300, Pahang, Malaysia.
*sorooshian@gmail.com

Abstract – *This article is in the form of a short tutorial discussion, presenting the logistic (logit) regression mathematical method. Logistic regression as one of the most popular modeling approach used in the research and decision makings, has its own specialty than other common linear regression method whereby the response variable is discrete and form of S-shaped. The paper starts with the short review of logistic regression method, formulation of the method and finally, authors' conclusion is the last section of this paper.*

Keywords: Logistic regression, Logit regression, Modeling, Regressor

1.0 INTRODUCTION

Same as our last data analysis tutorials (for example, [1]), the main objective of this paper is to revew Logistic regression (Logit regression) for new users. Logistic regression is practically applied for studying the relationship between categorical or qualitative outcome variables and one or more predictor variables [2]. Some of the mathematical concept defines logistic regression as logit regression or logit model. The name derives from the odds logarithm and based on the formulation of logistic function. Nicolis and Tondini [3] stated that this logistic regression has a binomial or multinomial distribution. Although, logistic regression very similar to linear regression model the response variable is discrete, the distribution of variables and residuals are non-normal and the relationship between response variable and the regressors is S-shaped than liner. Logit regression being chosen due to its successful employed in social science [2]. The main aim of logit regression is to find on the basis of sample data, the reliability of statistical model capable to identify the factors and variables underlying the economic growth of the enterprises [3]. According to Saha [2] stated that logit regression analysis used to analyze the different independent variables by calculating the probability of success over the probability of failure in the form of an odds ratio. Based on Westergren, Karkson, Anderson, Ohlsson and Hallberg [4] the odds ratio is a measure of effect size where describe the strength of association or non-independence between two binary data values. It is used as descriptive statistics unlike other measures of association for paired binary data like the relative scores which the odds ratio treats the two variable being compared symmetrically and can be estimated by using some type of non-random samples. There is some contradiction of hypothesis testing in logistic regression.

2.0 LOGIT REGRESSION

To analysis regressions, the following steps are listed [5]: A. Problem statement B. Variables Selection C. Collection of data D. Choice of fitting method E. Specification of the model F. Fitness and validation of the model In Logistic regression, the null hypothesis is that, the

Penerbit Akademia Baru
Journal of Advanced Research Design
ISSN (online): 2462-1943 | Vol. 21, No. 1. Pages 6-9, 2016

predictor coefficient is zero in the population. Hypothesis test tell whether there is sufficient evidence in the sample data to reject the null hypothesis and therefore to accept the alternative hypothesis that the predictor variable coefficient differ from zero. Confidence intervals can be applied for hypothesis testing as well as for regression coefficients. The odds ratio may be presented as:

$$L_{I=\left[\frac{P_I}{1-P_I}\right]} \quad (1)$$

$$P_I P_I = P_r[X_i = 1] = 1 - P_r[X_i = 0] \quad (2)$$

X_i is the independent variable corresponding to

$$i^{th} \text{ category for } i = 1,2, \ldots., n \quad (3)$$

Since the applied research is an application of scientific knowledge to understand the phenomenon and applicable to used therefore, logistic regression method will conducted in applied research by analyzing the probability of performing and non-performing. A logit regression is based on the formulation of the logistic function. Basically, logit is the natural logarithm (ln) of odds Y happening to probabilities (1-π) of Y not happening [6].

The logistic model,

$$\text{logit}(Y) = \ln\frac{\pi}{1-\pi} = \beta_o + X_1\beta_1 + X_2\beta_2 + \ldots \quad (4)$$

Thus,

$\pi = \text{Probability}$
$(Y = \text{outcome of interest} \mid X_1 = x_1 = x_2)$

$$Y = \frac{e^{\alpha+\beta_1 X_1+\beta_2 X_2}}{1+e^{\alpha+\beta x}} \quad (5a)$$

$$Y = \frac{1}{e^{-(\alpha+\beta_1 X_1+\beta_2 X_2)}} \quad (5b)$$

α = y-intercept
β = regression coefficient
e = 2.71828
x = independent variables

P will indicate as 1
if $\alpha + \beta x$ is not negative

P will indicate as 0
if $\alpha + \beta x$ is negative

As shown in the Figure 1, Logit regression is partially similar to the Classic linear regression; just logit regression's graph is based on using the natural logarithm of the odds of the regressors [7]. Moreover the figure presents S shape of logit regression graph.

Journal of Advanced Research Design
ISSN (online): 2462-1943 | Vol. 21, No. 1. Pages 6-9, 2016

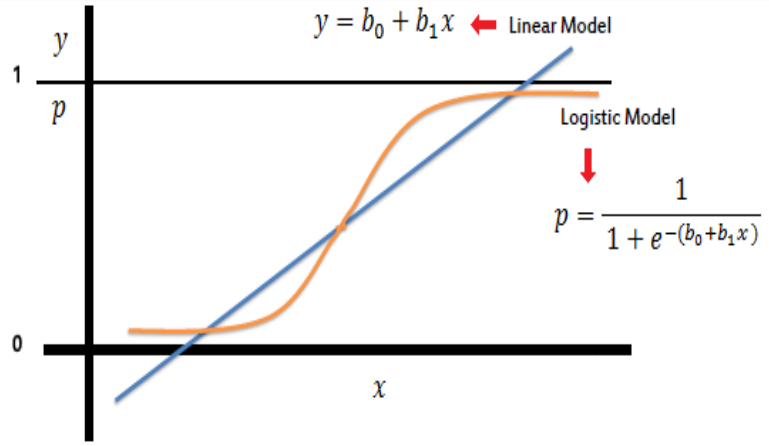

Figure 1: Difference between Classic linier regression and Logit regression [7].

3.0 CONCLUSION

Logistic Regression Method is widely used in the social science research as well as decision sciences. This method is emphasize the linear relationship between variables. Conceptually, it is similar to linear regression models, however it is a special type of generalized linear models which the response variable is discrete or categorical and has a binomial or multinomial distribution and the regressors is S-shaped, not a linear.

ACKNOWLEDGEMENT

The authors would like to thank faculty of industrial management, university Malaysia Pahang, for providing continuous support. The present research was made possible availing facility provided by GRS150359.

REFERENCES

[1] Aghelie, A., N. M. Mustapha , S. Sorooshian , N. A. Azizan. "Mathematical Modeling of Interrelationship Analysis to Determine Multi-Criteria Decision Making Casual Relations." Journal of Advanced Research Design 20, no.1 (2016): 18 – 33.

[2] Saha, Goutam. "Applying logistic regression model to the examination results data." Journal of Reliability and Statistical Studies 4, no. 2 (2011): 1-13.

[3] Nicolis, O., and G. Tondini. "Logit models for analysing and forecasting the performance of industrial enterprises in the Treviso area." Managerial Finance 32, no. 8 (2006): 654-672.

[4] Westergren, Albert, Siv Karlsson, Pia Andersson, Ola Ohlsson, and Ingalill R. Hallberg. "Eating difficulties, need for assisted eating, nutritional status and pressure ulcers in patients admitted for stroke rehabilitation." Journal of clinical nursing 10, no. 2 (2001): 257-269.

[5] Zulkifli, Norzima, Shahryar Sorooshian, and Alireza Anvari. "Modeling for Regressing Variables." Journal of Statistical and Econometric Methods 1, no. 2 (2012): 1-8.

Journal of Advanced Research Design
ISSN (online): 2462-1943 | Vol. 21, No. 1. Pages 6-9, 2016

[6] Peng, Chao-Ying Joanne, Kuk Lida Lee, and Gary M. Ingersoll. "An introduction to logistic regression analysis and reporting." The Journal of Educational Research 96, no. 1 (2002): 3-14.

[7] Logistic regression, http://www.saedsayad.com/logistic_regression.htm (access on: 10 June, 2015).