International Journal of Mathematics and Computational Science
Vol. 4, No. 3, 2018, pp. 79-85
http://www.aiscience.org/journal/ijmcs
ISSN: 2381-7011 (Print); ISSN: 2381-702X (Online)

On Estimation Methods for Binary Logistic Regression Model with Missing Values

Mohamed Reda Abonazel*, Mohamed Gamal Ibrahim

Department of Applied Statistics and Econometrics, Institute of Statistical Studies and Research, Cairo University, Cairo, Egypt

Abstract

This paper reviews some estimation methods for the binary logistic regression model with missing data in dependent and/or independent variables. Moreover, we present an empirical study for assessing the performance of these estimation methods under the existence of missing data. The results indicated that the regression imputation method is a very appropriate method for estimating the missing values in this model.

Keywords

EM Algorithm, Incomplete Data, Maximum Likelihood Estimation, Regression Imputation

Received: April 15, 2018 / Accepted: May 14, 2018 / Published online: June 14, 2018

1. Introduction

The use of logistic regression model dates back to 1845. It first appeared during the mathematical studies for the population growth at that time, see [10]. The term logistic regression analysis comes from logit transformation, which is applied to the dependent variable. This case, at the same time, causes certain differences both in estimation and interpretation. Logistic regression analysis is also called "binary logistic regression analysis", "multinomial logistic regression analysis" and "ordinal logistic regression analysis", this depending on the scale type and the number of categories of the dependent variable. Logistic regression is divided into two: "univariate logistic regression" and "multivariate logistic regression", see [22].

Logistic regression is widely used in many fields such as medical, business, economics, and so on. For example, in medical field suppose a patient has a disease (like HIV) based on the observed characteristics of patient (age, sex, various blood tests and urine tests). Another example, if you want to predict the election result for some national party or want to predict that whether voter will vote for congress or democratic party, based on the age, sex, income, caste, and many more characteristics, see [17].

Data related to confronted and researched cases in applied social sciences are mostly categorical (nominal) data with discrete value or data obtained by an ordinal scale. For instance, a man either works or unemployed; he is either a member of a group or not; the party in power is either from the right wing or the left wing; a student is either a graduate or not. In educational research, many problems relate with prediction of categorical results. For example, a student is either academically successful or not; he is either a slow learner or not; a teenager either has a tendency towards risky behavior or not. In literature of statistics, logistic or logit models are defined as regression models with categorical dependent variable.

Often, the dataset that used in our analysis is incomplete (includes missing values) in independent and/or dependent variables. In statistical literature, missing data procedure has been regarded as a probabilistic phenomenon. Little and Rubin [14] treated the missing values in the dataset as a set of random variables having a joint probability distribution. He developed a typology of missing data which became widely

* Corresponding author
E-mail address: mabonazel@hotmail.com (M. R. Abonazel)

practiced by researchers since then. The mechanisms consist of missing completely at random, missing at random, and missing not at random. These mechanisms define relationships between interesting variables and the likelihood of missing data. Each type of missing data dictates the performance of imputation techniques, see [1] and [9].

The researches that studying and handling the missing values in logistic regression models are very little until now, so this point still needs to more researches. Some of these researches; FitzGerald and Knuiman [7] examined a number of methods of handling missing outcomes in regressive logistic regression modelling of familial binary data. Consentino and Claeskens [4] derived explicit formulae for estimation in logistic regression models when some of the covariates are missing. Their approach allows for modelling the distribution of the missing covariates either as a multivariate normal or as a multivariate t-distribution. Sabbe et al [20] proposed an improved method that builds to handle missing data in both categorical and continuous predictors. Peng and Zhu [25] and Meeyai [24] compared several popular missing data handling methods in logistic regression model to study the performance of these methods in different situations. Maity et al [23] proposed a new method to improve the estimation of regression coefficients for this model. Their method based on penalizing the likelihood by multiplying it by a non-informative Jeffreys prior as a penalty term. They showed that this method reduces bias and is able to handle the issue of separation compared to the existing methods.

The aim of this paper is to review and study the performance of some estimation methods for the binary logistic regression model with missing data in dependent and/or independent variables, and then determinate the appropriate estimation method in this model.

The paper is organized as follows. Section 2 presents the model and maximum likelihood (ML) estimator. Section 3 discusses the different methods to handle missing data. In section 4, an empirical study has been presented for assessing the performance of different estimation methods under the existence of missing data. Finally, section 5 offers the concluding remarks.

2. Logistic Regression Models

Logistic regression measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function, which is the cumulative logistic distribution. Logistic regression can be seen as a special case of the generalized linear model (GLM) and thus similar to linear regression. In particular, the key differences between these two models can be seen in the following two features of logistic regression. First, the conditional distribution y | x is a Bernoulli distribution rather than a Gaussian distribution, because the dependent variable is binary. Second, the predicted values are probabilities and are therefore restricted to (0,1) through the logistic distribution function because logistic regression predicts the probability of particular outcomes.

Since the independent variables in the model are categorical, or a mix of continuous and categorical while the dependent variable is categorical, so logistic regression analysis is necessary. Also, since the dependent variable is dichotomous one cannot predict a numerical value for it using logistic regression, so the usual regression least squares deviations criteria for best fit approach of minimizing error around the line of best fit is inappropriate. Instead, logistic regression employs binomial probability theory in which there are only two values to predict: that probability is 1 rather than 0, i.e. the event/person belongs to one group rather than the other. Logistic regression forms a best fitting equation or function using the ML method, which maximizes the probability of classifying the observed data into the appropriate category given the regression coefficients, see [3].

2.1. The Binary Model

Generally, logistic regression is generally thought of as a method for modeling in situations for which there is a binary response variable and the predictor variables can be numerical or categorical (including binary). Also, logistic regression can be used when there are more than two possible outcomes for the response. But here the focus will be in the typical binary response version. Figure 1 shows the difference between linear regression and the binary logistic regression.

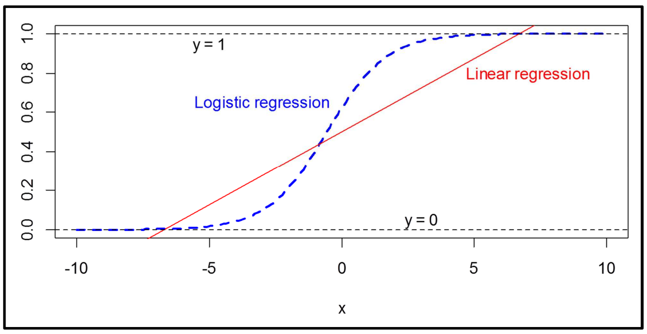

Figure 1. Linear versus logistic regression.

Letting y be the binary response variable, it is assumed that $p(y=1)$ is possibly dependent on $\bar{X}$, a vector of predictor values. The goal is:

$$p(\bar{X}) = p(y = 1|\bar{X}). \tag{1}$$

If the model $p(\bar{X})$ as a linear function of predictor variables, e.g., $\beta_o + \beta_1 x_1 + \cdots + \beta_K x_K$ then the fitted model can result in estimated probabilities which are outside of [0,1]. What tends to work better is to it's called multiple logistic regression

$$p(\bar{X}) = 1/\left[1 + e^{-(\beta_o + \beta_1 x_1 + \cdots + \beta_K x_K)}\right]. \quad (2)$$

The outcome of the regression is not a prediction of a y value, as in linear regression, but a probability of belonging to one of two conditions of y, which can take on any value between 0 and 1 rather than just 0 and 1. Unfortunately a further mathematical transformation (a log transformation) is needed to normalize the distribution. This log transformation of the p values to a log distribution enables us to create a link with the normal regression equation. The log distribution (or logistic transformation of p) is also called the logit of p or $logit\ (p)$. $Logit\ (p)$ is the log to base e of the *odds* ratio or likelihood ratio that the dependent variable is 1. In symbols it is defined as:

$$Logit\ (odds) = log\left(\frac{p(\bar{X})}{1-p(\bar{X})}\right) = ln\left(\frac{p}{1-p}\right). \quad (3)$$

Where p can only range from 0 to 1, logit (p) scale ranges from negative infinity to positive infinity and is symmetrical around the logit of 0.5 (which is zero). The form of the logistic regression equation is:

$$Logit\ (odds) = ln\left(\frac{p}{1-p}\right) = \beta_o + \beta_1 x_1 + \cdots + \beta_K x_K. \quad (4)$$

This looks just like a linear regression and although logistic regression finds a 'best fitting' equation, just as linear regression does, the principles on which it does so are rather different. Instead of using a least-squared deviations criterion for the best fit, it uses a ML method, which maximizes the probability of getting the observed results given the fitted regression coefficients. A consequence of this is that the goodness of fit and overall significance statistics used in logistic regression is different from those used in linear regression. p can be calculated with the following formula

$$p = \left(e^{\beta_o + \beta_1 x_1 + \cdots + \beta_K x_K}\right)/\left(1 + e^{\beta_o + \beta_1 x_1 + \cdots + \beta_K x_K}\right), \quad (5)$$

where p is the probability that a case is in a particular category, e is the base of natural logarithms (approx. = 2.72), β_0 is the constant of the equation, and $\beta_k (k = 1, 2, \ldots, K)$ are the coefficients of the predictor variables.

2.2. Logistic Regression Assumptions

1. Logistic regression does not assume a linear relationship between the dependent and independent variables.
2. The dependent variable must be a dichotomy (2 categories).
3. The independent variables need not be interval, no normally distributed, no linearly related, no of equal variance within each group.
4. The categories (groups) must be mutually exclusive and exhaustive; a case can only be in one group and every case must be a member of one of the groups.

2.3. Logistic Transformation

To achieve this, a regression is first performed with a transformed value of Y, called the "Logit function":

$$Logit(Y) = ln\ (odds) = \beta_o + \beta_1 x_1 + \cdots + \beta_K x_K, \quad (6)$$

where "*odds*" refers to the odds of Y being equal to 1. In other words, "*odds*" is defined as the probability of belonging to one group divided by the probability of belonging to the other: $odds = p/\,(1-p) \rightarrow [0, \infty[$.

This means that the values of $odds$ are always positive. But the log ($odds$) are continuous: $\log{(odds)} = \ln\left(\frac{p}{1-p}\right) \rightarrow (-\infty, \infty)$. This equation can be rewritten in terms of probability p as: $p = odds/(1 + odds)$.

Table 1. "Logit" transformation of the probability.

Measure	Min	Max	Name
$P(Y=1)$	0	1	probability
$\frac{P(Y=1)}{1-P(Y=1)}$	0	∞	odds
$\text{Log}\left[\frac{P(Y=1)}{1-P(Y=1)}\right]$	$-\infty$	∞	log-odds or logit

2.4. Maximum Likelihood Estimation

Although you will probably use a statistical package to compute the estimates, here is a brief description of the underlying procedure. Because logistic regression predicts probabilities, rather than just classes, one can fit it using likelihood. For each training data-point, we have a vector of features, x_i, and an observed class, y_i. The probability of that class was either if $y_i = 1$, or $(1 - p)$ if $y_i = 0$. The likelihood function for simple logistic regression is

$$L(\beta_0, \beta_1) = L = \prod_{i=1}^{n} p(\,x_i)^{y_i}\ [1 - p(x_i)]^{1-y_i}. \quad (7)$$

The log-likelihood turns products into sums:

$$\log(L) = \sum_{i=1}^{n} y_i \log p(x_i) + (1 - y_i)\log[1 - p(x_i)]$$

$$= \sum_{i=1}^{n} -\log(1) + e^{(\beta_O + x_i\beta_1)} + \sum_{i=1}^{n} y_i\,(\beta_0 + x_i\beta_1). \quad (8)$$

Typically, to find the ML estimates we'd differentiate the log likelihood with respect to the parameters and set the derivatives equal to zero to get the estimates. Since this equation is nonlinear in β, some special methods should be used in order to obtain the estimated parameters. The iteratively re-weighted least squares (IRLS) method can be applied to get the solutions. The ML estimator of β can be obtained by using IRLS algorithm as follows:

$$\hat{\beta}_{MLE} = (X'\widehat{W}X)^{-1}X'\widehat{W}\hat{Z}, \quad (9)$$

where $\widehat{W} = diag\{\hat{P}_i(1-\hat{P}_i)\}$ and $\hat{z}_i = \log(\hat{P}_i) + \frac{y_i - \hat{P}_i}{\hat{P}_i(1-\hat{P}_i)}$ is the i^{th} element of the vector $\hat{Z}$. The hats in the equations show the iterative process.

3. Incomplete Data

To decide how to handle missing data, it is helpful to know why they are missing. There are three general missingness mechanisms, moving from the simplest to the most general.

3.1. Missingness Mechanisms

3.1.1. Missingness Completely at Random (MCAR)

A variable is missing completely at random if the probability of missingness is the same for all units, for example, if each survey respondent decides whether to answer the "earnings" question by rolling a dice and refusing to answer if a "6" shows up. If data are missing completely at random, then throwing out cases with missing data does not bias your inferences, see [9].

3.1.2. Missingness at Random (MAR)

Most missingness is not completely at random, as can be seen from the data themselves. For example, the different nonresponse rates for whites and blacks indicate that the "earnings" question in the Social Indicators Survey is not missing completely at random. A more general assumption, missing at random, is that the probability a variable is missing depends only on available information. Thus, if sex, race, education, and age are recorded for all the people in the survey, then "earnings" is missing at random if the probability of nonresponse to this question depends only on these other, fully recorded variables, see [9].

3.1.3. Missingness Not at Random (MNAR)

Missingness is no longer "at random" if it depends on information that has not been recorded and this information also predicts the missing values. For example, suppose that "surly" people are less likely to respond to the earnings question, surliness is predictive of earnings, and "surliness" is unobserved. Or, suppose that people with college degrees are less likely to reveal their earnings, having a college degree is predictive of earnings, and there is also some nonresponse to the education question. Then, once again, earnings are not missing at random, see [9].

3.2. Estimation Methods for Incomplete Data

3.2.1. Deletion-Based Methods: Listwise and Pairwise Deletion

Listwise and pairwise deletion techniques are the most common techniques to handling missing data in regression models, see [19]. An important assumption to using either of these techniques is the data is MCAR. In other words, the researcher needs to support that the probability of missing data on their dependent variable is unrelated to other independent variables as well as the dependent variable itself.

Listwise deletion technique (complete-case analysis) removes all data for a case that has one or more missing values. This technique is commonly used if the researcher is conducting a treatment study and wants to compare a completers analysis (listwise deletion) vs. an intent-to-treat analysis (includes cases with missing data imputed or taken into account via a algorithmic method) in a treatment design. In most other cases, it is often disadvantageous to use listwise deletion technique because the assumptions of MCAR are typically rare to support. Because of this, listwise deletion technique produces bias parameters and the estimates.

Pairwise deletion technique (available-case analysis) attempts to minimize the loss that occurs in listwise deletion technique. An easy way to think of how pairwise deletion works is to think of a correlation matrix. A correlation measures the strength of the relationship between two variables. For each pair of variables for which data is available, the correlation coefficient will take that data into account. So, pairwise deletion technique maximizes all data available by an analysis by analysis basis. Strength to this technique is that it increases power in your analyses. Though this technique is typically preferred over listwise deletion, it also assumes that the missing data are MCAR. A disadvantage with the use of pairwise deletion is that the standard of errors computed by most software packages uses the average sample size across analyses. This tends to produced standard of errors that are underestimated or overestimated. Researchers have also associated pairwise deletion as a source for non-positive definite matrices in multivariate and contemporary statistical analyses, such as Structural Equation Modeling[1], see [13].

3.2.2. Expectation-Maximization (EM) Algorithm

A general approach to getting ML estimates with missing data (Two-step procedure):

Step (1): Expectation (E): Finding the expected value of the log-likelihood for the observed data, based on current

1 For details about contemporary statistical analysis and structural equation modeling, see for example [2], [5], and [6].

parameter values.

Step (2): Maximization (M): Maximizing the expected likelihood to get new parameter estimates.

The EM algorithm has been a popular technique for obtaining ML estimators in GLMs with missing covariate data. A good book on the theory and applications of EM is [16]. Fuchs [8] used the EM algorithm to get ML estimators for log-linear models with incomplete data. Little and Schluchter [15] used the EM algorithm to obtain estimates in a regression model with ignorable missing categorical and continuous covariates. Schluchter and Jackson [21] used EM to find parameter estimates in log-linear models with ignorable missing covariate data.

A general method for estimation in the presence of missing covariates has been proposed by Ibrahim [11], who used EM by the method of weights to find the ML estimators. Ibrahim's method applies to any parametric regression model, including GLMs, nonlinear models, random-effects models, frailty models, and parametric and semi parametric survival models.

3.2.3. Regression Imputation

Regression imputation is useful for imputing continuous variables. If Y represents the continuous variable with missing values to be imputed and X represents a vector of predictor variables, then a linear regression model is fit (usually) by the method of ordinary least squares (OLS). All of the usual assumptions concerning OLS regression apply. The idea is to estimate the regression coefficients and the error variance for a regression model. The errors are assumed to be normally distributed so that the regression coefficients and the error variance have known statistical distributions. Imputations are generated on the basis of predictions generated by random draws from the statistical distributions for the coefficients and the sampling variance. One common issue with OLS regression is that "impossible" predictions are possible.

For example, a regression for total hospital charges might produce imputations with negative values for some observations. This might not be of great concern if the objective is simply to estimate average charges over a large sample. The mean might still be unbiased. However, in the event that negative imputations are problematic, the analyst might choose to fit a log-linear model to ensure positive predictions. Other solutions replace negative values with "minimum" positive values. However, that procedure will most likely bias the estimates. For more details about these procedures, see [16].

4. Application

In this section, a case study on a pharmaceutical firm that developed a particular drug for women has been presented. The aim is to understand the characteristics that cause some of them to have an adverse reaction to a particular drug. The data collected from 15 women who had such a reaction, and 15 who did not. The independent variables that have been measured are: Systolic Blood Pressure (BP), Cholesterol Level, Age of the women (Age), Whether or not the woman was pregnant (1 = yes). While the dependent variable indicates if there was an adverse reaction (1 = yes). This data studied by [18].

4.1. Complete Data Analysis

Some basic descriptive statistics and ML estimates of the model are given in the following tables:

Table 2. Some descriptive statistics for the data.

Variable	Mean	Std. Deviation	Min	Max
Dependent				
Drug Reaction (y)	.50	.509	0	1
Independents				
BP	127.33	22.846	95	180
Cholesterol	185.07	28.463	130	250
Age	37.77	18.796	16	81
Pregnant	.50	.509	0	1

Table 3. ML estimates of the model in case of complete data.

Parameter	Estimates	Std. Error	Exp (B)	Wald	P-Value
Intercept	17.874	10.1585	.000	3.096	.078
BP	.018	.0268	.982	.463	.496
Cholesterol	-.027	.0246	. 1.027	1.182	.277
Age	-.265	.1142	4918.147	5.404	.020
Pregnant	-8.501	3.8842	1.304	4.790	.029

Table 3 shows that "Age" and "Pregnant" are statistically significant because the P-values for their variables less than 0.05, while "BP" and "Cholesterol" are not significant. Moreover, the deviance R-Squared value is 37.87%. And the estimated equation of the model is:

$$\begin{aligned} Log\ (odds\ of\ drug\ reaction) &= 17.874 + .018\, BP \\ &- 0.027\, Cholesterol \\ &- 0.265\, Age - 8.501\, Pregnant. \end{aligned} \quad (10)$$

To check the goodness-of-fit of the model, the following hypothesis must be testing:

H_0: The data are consistent with a specified distribution.

Table 4. Goodness of Fit test.

Goodness of Fit	Value	D. f	Chi-Square	P-Value
Deviance	21.841	25	21.84	0.645
Pearson chi sq.	20.010	25	20.01	

Table 4 shows that the P-value is large than 0.05, then the model is very suitable for data.

After doing the regression analysis on the complete data, the results indicated that the parameters of blood pressure and cholesterol are not significant, so it has no effect on Drug Reaction, because both blood pressure and cholesterol are increasing naturally in pregnant women according to the medical diagnosis, and thus it has no effect on drug reaction. But the other parameters of both age and pregnant are statistically significant, which means that it has a significant on drug reaction by the value of the estimated parameter.

4.2. Incomplete Data Analysis

Assuming that some of our data are missing by approximately 10% and then estimate these data and doing the same analysis. To estimate these missing values, EM and regression imputation methods have been used. To compare between these two methods, two criterions have been used; R-square and Akaki information criterion (AIC). The missing values in all variables have been generated as follows:

A. Dependent variable.

B. Independent variables.

C. Dependent and independent variables together.

After the missing values have been generated in the data by random, little's [12] test has been used to check that missing data biased to MCAR. The null hypothesis for this test is:

H_0: Data is missing completely at random

Table 5. Little's MCAR test.

	Dependent variable	Independent variables	Dependent and independent variables
Chi-square	2.209	9.069	12.278
P-value	.530	.170	.056

Note from Table 5 that all P-values more than 0.05 (Accept H_0) for the three cases. This means that the missing data that generated is completely at random.

4.3. Comparison Between Estimation Methods

Table 6. Goodness of fit measures for estimation methods.

Complete / Incomplete	Estimation Method	R-Square	AIC
I. Complete Data	ML	.643	31.84
II. Incomplete Data			
a. Dependent variable	EM	.357	42.23
	Regression	.683	29.98
b. Independent variables	EM	.596	33.75
	Regression	.690	29.72
c. Dependent and independent variables	EM	.400	40.87
	Regression	.566	34.98

Table 6 shows that the regression imputation method has lower AIC and higher R-square than EM method. Moreover, the AIC and R-square values of regression imputation method are very close to ML method. So, we can conclude that, in our study, the regression imputation is the best method for estimating the missing values in this model.

5. Conclusion

This paper reviewed some estimation methods for the binary logistic regression model with missing data in dependent and/or independent variables. Moreover, we presented an empirical study for assessing the performance of different estimation methods under the existence of missing data. According to the results, it was found that the regression imputation is a very appropriate method for estimating the missing values in this model.

In future work, the Monte Carlo simulation study[2] can be performed to compare the different estimation methods for missing values in different situations (a different samples sizes, a different number of independent variables, and so on).

References

[1] Allison, P. D. (2002). Missing data: Quantitative applications in the social sciences. *British Journal of Mathematical and Statistical Psychology*, *55* (1), 193-196.

[2] Byrne, B. M. (2013). Structural equation modeling with AMOS: Basic concepts, applications, and programming. Routledge.

[3] Burns, R. P., & Burns, R. (2008). *Business research methods and statistics using SPSS*. Sage.

[4] Consentino, F., & Claeskens, G. (2011). Missing covariates in logistic regression, estimation and distribution selection. *Statistical Modelling*, *11* (2), 159-183.

[5] El-Sheikh, A. A., Abonazel, M. R., & Gamil, N. (2017). A Review of Software Packages for Structural Equation Modeling: A Comparative Study. *Applied Mathematics and Physics*, *5* (3), 85-94.

[6] El-Sheikh, A. A., Abonazel, M. R., & Gamil, Noha (2017). A review of estimation methods for structural equation modeling. Working paper. Institute of Statistical Studies and Research. Cairo University, Egypt.

[7] FitzGerald, P. E., & Knuiman, M. W. (1998). Theory and Methods: Estimation in Regressive Logistic Regression Analyses of Familial Data with Missing Outcomes. *Australian & New Zealand Journal of Statistics*, *40* (3), 305-316.

[8] Fuchs, C. (1982). Maximum likelihood estimation and model selection in contingency tables with missing data. *Journal of the American Statistical Association*, *77* (378), 270-278.

[9] Houchens, R. (2015). *Missing Data Methods for the NIS and the SID*. HCUP Methods Series Report # 2015-01. Agency for Healthcare Research and Quality [accessed on June 22, 2015]. Available: https://www.hcup-us.ahrq.gov/reports/methods/2015_01.pdf

2 Abonazel [26] provided a practical guide to conduct Monte Carlo simulation studies using R language.

[10] Hsieh, F. Y., Bloch, D. A., & Larsen, M. D. (1998). A simple method of sample size calculation for linear and logistic regression. *Statistics in medicine*, *17* (14), 1623-1634.

[11] Ibrahim, J. G. (1990). Incomplete data in generalized linear models. *Journal of the American Statistical Association*, *85* (411), 765-769.

[12] Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83* (404), 1198-1202.

[13] Little, R. J. (1992). Regression with missing X's: a review. *Journal of the American Statistical Association, 87*, 1227-1237.

[14] Little, R. J., & Rubin, D. B. (2002). Statistical Analysis with Missing Data. John Wiley & Sons.

[15] Little, R. J., & Schluchter, M. D. (1985). Maximum likelihood estimation for mixed continuous and categorical data with missing values. *Biometrika*, *72* (3), 497-512.

[16] McLachlan, G. J., & Krishnan, T. (1997). Wiley series in probability and statistics. *The EM Algorithm and Extensions, Second Edition*, 361-369.

[17] Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European sociological review*, *26* (1), 67-82.

[18] Nargundkar, S. (2015). Chapter 12: Logistic Regression for Classification and Prediction. https://www.coursehero.com/file/7073839/Logistic-Regression.

[19] Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74,* 525-556.

[20] Sabbe, N., Thas, O., & Ottoy, J. P. (2013). EMLasso: logistic lasso with missing data. *Statistics in medicine*, *32* (18), 3143-3157.

[21] Schluchter, M. D., & Jackson, K. L. (1989). Log-linear analysis of censored survival data with partially observed covariates. *Journal of the American Statistical Association*, *84* (405), 42-52.

[22] Stephenson, B., Cook, D., Dixon, P., Duckworth, W., Kaiser, M., Koehler, K., & Meeker, W. (2008). Binary response and logistic regression analysis. Available: http://www.stat.wisc.edu/~mchung/teaching/MIA/reading/GLM.logistic.Rpackage.pdf

[23] Maity, A. K., Pradhan, V., & Das, U. (2017). Bias Reduction in Logistic Regression with Missing Responses when the Missing Data Mechanism is Nonignorable. *The American Statistician*, (just-accepted).

[24] Meeyai, S. (2016). Logistic Regression with Missing Data: A Comparison of Handling Methods, and Effects of Percent Missing *Values. Journal of Traffic and Logistics Engineering*, 4 (2).

[25] Peng, C. Y. J., & Zhu, J. (2008). Comparison of two approaches for handling missing covariates in logistic regression. *Educational and Psychological Measurement*, *68* (1), 58-77.

[26] Abonazel, M. R. (2018). A practical guide for creating Monte Carlo simulation studies using R. *International Journal of Mathematics and Computational Science*, *4* (1), 18-33.