Self-Selecting Robust Logistic Regression Model

Idelphonse Léandre Tawanou Gbohounme[1], Oscar Owino Ngesa[2] & Jude Eggoh[3]

[1] Pan African University (PAUISTI), Institute for Basic Sciences Technology and Innovation, Kenya

[2] Taita Taveta University College, Kenya

[3] University of Angers, France

Correspondence: Idelphonse Léandre Tawanou Gbohounme, Pan African University (PAUISTI), Institute for Basic Sciences Technology and Innovation, Kenya. Tel: 337-585-521-64. E-mail: gitawanou@gmail.com

Abstract

Logistic regression model is the most common model used for the analysis of binary data. However, the problem of atypical observations in the data has an unduly effect on the parameter estimates. Many researchers have developed robust statistical model to solve this problem of outliers. Gelman (2004) proposed GRLR, a robust model by trimming the probability of success in LR. The trimming values in this model were fixed and the user is required to specify this value well in advance. In particular this study developed SsRLR model by allowing the data itself to select the alpha value. We proposed a Restricted LR model to substitute the LR in presence of outliers. We proved that the SsRLR model is the more robust to the presence of leverage points in the data. Parameter estimations is done using a full Bayesian approach implemented in WinBUGS 14 software.

Keywords: logistic regression, robust model, leverage points, Bayesian approach, trimming approach

1. Introduction

Many dependent variables of interest in the social sciences are usually not a continuous variables. In most cases, the outcomes are categorical with two levels, namely, yes/no, success/failure, 0/1. Such variables are called binary responses or dichotomous. Binary logistic regression is a helpful way of explaining the relationship between one or more independent variables and a binary response. By assuming a binary variable y with $\pi = Pr(y = 1)$ the probability of success, the classical LR model is defined as:

$$LR : \qquad \pi = logit^{-1}(X^T\beta) \tag{1}$$

where X is a vector of p independent variables and β is a p dimensional vector of regression coefficients for the predictor variables.

Robustness is a subject highly developed in the fields of estimation of the position and scale of simple and multiple regression. Attention has been paid to the robust logistic regression, which is an area where outliers may also appear. Pregibon (1981) started by developing an analytical measure to assist in the detection of outliers and leverage points and quantify their effect on diverse aspects of the maximum likelihood fit. Thereafter, a good number of robust estimation procedures in the context of logistic regression have been examined. Gelman (2004) proposed a GRLR using a trimming approach. The approach used a trimming value α, 0.01 chance of random error in both direction of the interval [0,1]. Gelman's Robust Logistic Regression model is defined as:

$$GRLR : \qquad \pi = 0.01 + 0.98logit^{-1}(X^T\beta) \tag{2}$$

where 0.01 and 0.98 are fixed. That model requires the statistician to specify these values beforehand. SsRLR solves this problem in the GRLR model by relaxing this restriction and letting these probabilities to be self selected by the data at hand so that only a prior distribution, say Uniform $[a, b]$ with a and b belonging to [0,1] is given.

2. Methods

2.1 Robust Logistic Regression

Robust signifies remaining resistant against some irregular deviations. In statistics, models are a simple estimation of reality. The models that underlie numerous statistical process are very optimistic and in real data, big errors happen with unpredictable large frequency. An observation that lies an abnormal distant from the mass of data is set as an outlier. Robustness means insensitivity against some divergence from the right model. Robust process was initiated in the works of Tukey (1960) and further, formal models of robustness have been expanded in 1970's.

In regression models, the purpose of robust methods is to detect the outliers and extremely influential data points, leverage points, and to end by describing the goodness of fit for the data. One of the first result works related to Least Square, as a robust estimator, was carried out by Edgeworth (1887), who enhanced the proposal of Boscovich (1757) (Koenker & Bassett. 1985). This estimator is the least absolute deviation (LAD).

Logistic regression is concerned with explaining the probability of a specific response in terms of a number of regressors using a sample of relevant data. Pregibon (1981) affirmed that the estimated LR correlation may be extremely influenced by outliers. Dealing with outliers necessitates the use of robust logistic regression models to overcome their influence on the LR model. Researches in this direction have been conducted by Hubert (1973), Pregibon (1981), Rousseeuw et Al. (1987), Yohai (1987), Copas (1988) and Rousseeuw (2003).

Trimming is an extensive approach to robustifying of statistical process. It permits one to detect outliers and eliminate them from the data exploited in the estimation procedure. Trimming has been expanded highly by different authors in least squares regression, multivariate analysis and other areas (Rousseeuw (1984), Rousseeuw & Van Driessen (1999), where additional mentions can be obtained). It appears attractive to apply trimming also in logistic regression to find outliers and to control their influences.

On the other hand, the outlier can disturb statistical models and results in an expected model differ significantly from the exact one. Outliers in LR may occur in the Y-space called misclassification-type error (Copas, 1988), the X-space considered as leverage points or in both spaces. Outlying cases in this study are only based on the covariate corruptions. The robust LR model in this study is based on that approach of trimming probability whose estimation procedure is related to Bayesian inference using Gibbs sampler and Metropolis-Hastings Algorithm.

2.2 Proposed SsRLR Model

In this work, we improve the model of Andrew Gelman (2004) by developing a self-selecting robust logistic regression. Suppose $y = (y_1, y_2, \ldots, y_n)$ are n independent observations where y_i are binary responses data defined as:

$$y_i = \begin{cases} 1 & \text{if success} \\ 0 & \text{otherwise} \end{cases}$$

Binary regression models assume that $y_i \sim \text{Ber}(\pi_i)$ with $\pi_i = Pr(y_i = 1)$ the probability of success for each observation. From that, the robust model we are developing is as follows:

$$\pi_i = \alpha + (1 - 2\alpha)logit^{-1}(X^T\beta) \tag{3}$$

where X is a vector of p independent variables and β is a p dimensional vector of regression coefficients for the predictor variables.

As opposed to other studies where the value of α is set beforehand by the statistician, we allow this to be determined from the data itself. In particular since we are working in the Bayesian paradigm, we give this value α a uniform prior distribution.

2.3 Estimation using Bayesian Approach

Bayesian approach in estimation is used to minimize risk estimation and to obtain the optimal estimates. To proceed, we follow the usual pattern for all Bayesian analyses by writing down the likelihood function of the data, forming a prior distribution over all unknown parameters and using Bayes theorem to find the posterior distribution over all parameters.

Likelihood function

In particular, once the probability of success depending on the covariates is obtained, the likelihood function is:

$$L(\beta, \alpha | X, y) = \prod_{i=1}^{n} [(\pi_i)^{y_i} (1 - \pi_i)^{1-y_i}] \tag{4}$$

where π_i represents the probability of success and y_i the binary responses data. In our model we have:

$$\pi_i = \alpha + (1 - 2\alpha) \frac{e^{\beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip}}} \tag{5}$$

Hence the likelihood function of the binary responses data of n independent observations is:

$$L(\beta, \alpha, \sigma | X, y) = \prod_{i=1}^{n} A_i B_i \tag{6}$$

where:

$$A_i = \left[\alpha + (1 - 2\alpha) \frac{e^{\beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip}}} \right]^{y_i}$$

$$B_i = \left[1 - \left(\alpha + (1 - 2\alpha) \frac{e^{\beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip}}} \right) \right]^{1 - y_i}.$$

2.3.1 Prior Distributions

In this analysis, a non informative normal prior was assigned to the regression coefficients β, The parameter under study α is given a uniform prior distribution U[a,b].

2.3.2 Posterior Distributions

To derive the posterior distribution, we multiply the prior distribution over all parameters by the likelihood function. Thus we have:

$$P_{post}(\theta|X, y) = L(\beta_j, \alpha, \sigma|X, y) P_{pri}(\beta_j) P_{pri}(\alpha) \tag{7}$$

where

$$P_{pri}(\beta_j) = \prod_{j=0}^{p} \frac{1}{\sqrt{2\pi}\sigma_j} exp\left[-\frac{1}{2} \left(\frac{\beta_j}{\sigma_j} \right)^2 \right] \tag{8}$$

$$P_{pri}(\alpha) = \frac{1}{b - a}, a \leq \alpha \leq b \tag{9}$$

are β and α prior distributions respectively.

Most often, Bayes estimators of θ cannot be computed explicitly and we have to look for Monte Carlo simulation method, using Gibbs sampler algorithm where the computing of Bayes estimators does not pose great difficulty. For each model, we ran 10,000 Markov chain Monte Carlo (MCMC) iterations, with the initial 1,000 discarded to cater for the burn-in period and thereafter keeping every tenth sample value. MCMC convergence of all models parameters were accessed by checking trace plots and auto correlation plots of the MCMC output.

2.4 Gibbs Sampler Algorithm

Consider the basic case: f(x,y). Assume f((x|y)) and f(y|x) available. We can then generate what one will call Gibbs sequence as follows: starting from a value x_0, y_0 is generated with $\pi(.|x_0)$, then x_1 with $\pi(.|y_0)$, and y_1 with $\pi(.|x_1)$ and so on.

After M iterations of this scheme, it comes a sequence $(x_0, y_0, x_1, y_1, \ldots, x_M, y_M)$. For M large enough, x_M is a realization of X.

In the Bayesian framework, the Gibbs algorithm (Geman. 1984) will allow to obtain a realization of the parameter $\theta = (\theta_1, \ldots, \theta_m)$ following the posterior distribution $\pi(\theta|x)$ as soon as one is capable of expressing the conditional distributions: $\pi(\theta_i|\theta_j; x)$, $j \neq i$. Thus, Gibbs sampling consists of:

Starting from an initial vector $\theta^{(0)} = (\theta_1^{(0)}, \ldots, \theta_m^{(0)})$. At the $(p + 1)^{th}$ step, with the vector $\theta^{(p)} = (\theta_1^{(p)}, \ldots, \theta_m^{(p)})$, simulating

$$\begin{aligned}
\theta_1^{(p+1)} &= \pi(\theta_1|\theta_2^{(p)}, \theta_2^{(p)}, \ldots, \theta_m^{(p)}; x) \\
\theta_2^{(p+1)} &= \pi(\theta_2|\theta_1^{(p+1)}, \theta_3^{(p)}, \ldots \theta_m^{(p)}; x) \\
&\ldots \\
\theta_m^{(p+1)} &= \pi(\theta_m|\theta_1^{(p+1)}, \theta_2^{(p)}, \ldots, \theta_{m-1}^{(p)}; x)
\end{aligned}$$

$$\tag{10}$$

Successive iterations of this algorithm successively generate the states of a Markov chain $\{\theta^p, p > 0\}$ for values $\aleph^{\otimes m}$. The transition probability from θ' to θ is expressed as:

$$K(\theta', \theta) = K_1(\theta', \theta) \times K_2(\theta', \theta) \tag{11}$$

where:

$$
\begin{aligned}
K_1(\theta', \theta) &= \pi(\theta_1|\theta'_2, \ldots, \theta'_m) \times \pi(\theta_2|\theta_1, \theta'_3, \ldots, \theta'_m) \\
K_2(\theta', \theta) &= \pi(\theta_3|\theta_1, \theta_2, \theta'_4, \ldots, \theta'_m) \times \ldots \times \pi(\theta_m|\theta_1, \ldots, \theta_{m-1}).
\end{aligned}
\tag{12}
$$

This shows that the chain admits an invariant measure which is the posterior. For a sufficiently large number of iterations, the vector θ thus obtained may be considered as a realization of the posterior.

2.5 Model Diagnostics

The models goodness of fit were compared using the Deviance Information Criterion (DIC) as suggested by Spiegelhalter (2002). The best fitting model is one with the smallest DIC. The DIC value is given by $DIC = \overline{D}(\theta) + pD$, where $\overline{D}$ is the posterior mean of the deviance that measures the goodness of fit, and pD gives the effective number of parameters in the model which penalizes for complexity of the model. However, several authors have stated that a difference in DIC of 3 between two models can not be distinguished while a difference between 3 and 7 can be weakly differentiated.

For further model assessment, we associated the Bayesian Information Criterion (BIC). In statistic, the Bayesian information criterion or Schwarz criterion is a criterion for model selection among a finite set of models and the model with the lowest BIC is preferred (Schwarz. 1978). It is based, in part, on the likelihood function and it is closely related to the Akaike information criterion (AIC). BIC value is given by $BIC = Dhat + 2p * log(n)$ where $Dhat = -2logL(\theta^*|y)$ with $L(\theta^*|y)$, the likelihood of each model, p the number of parameters and n the sample size.

3. Simulation Study

We carried out a simulation study to investigate the robustness of the three models namely: the Self-Selecting Robust Logistic Regression (SsRLR) model, Gelman's Robust Logistic Regression (GRLR) model and the ordinary Logistic Regression (LR) model. Following the simulation study carried out by Croux & Haesbroeck. (2003), LR model is generated with two independent normally distributed covariates. The additive noise ε_i is selected from a logistic distribution defined as:

$$
logit(\pi_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + (\varepsilon_i \geq 0)
\tag{13}
$$

The true parameter values are $\beta = (0, 2, 2)$ with sample size $n = 200$. The study was based under a variety of situation. First, data without contamination was taken with two independent normally distributed covariates with zero mean and unit variance. Second, to examine the robust properties of all models, we introduced outliers by contaminating the data similarly to the idea proposed by Victori (2002). We generated the outliers in R software by corrupting the covariates. This consists of randomly choosing a certain t proportion $(3\%, 5\%, 7\%)$ from both covariates and replace them with a sample X_i chosen from $N(t, 10, 2)$. The response variable for each proportion was then generated from the new corrupted covariates. Finally the generated binary response data was contaminated under different percentages of leverage points. Thereafter, the three logistic models were applied to these data generated. In order to better handle those outliers, our robust model proposed to the contaminated binary data response itself to select the value of the probability α. After getting that significant alpha value for the robust model, we compared the goodness of fit of the three logistic regression models.

4. Results

For each simulated data set, we estimated and recorded the parameters β and α. In particular we focus on investigating how much each model performs in presence of outliers in the binary response data. In assessing that performance, we compute and compare their DIC and BIC.

4.1 Model Assessment and Comparison

The first finding involved the LR model. In fact, the generated outliers values between 5 and 10 caused the LR not to run, giving "Trap Message" and no output while the SsRLR model model takes care of the leverage points without any problem. We got output and summary statistics by using a Restricted Logistic Regression (RLR) model where y_i $\sim$Bern(π_1) is defined as:

$$
RLR : logit(\pi) = \beta_0 + \beta_1 X_1 + \beta_2 X_2
\tag{14}
$$

π_1 = min(1,max(0.001,π))

It can be deduced that fitting ordinary logistic regression with outliers can get "Trap Message" and no output without using Minmax in WinBUGS. Figures show a visual representation of the distribution of the data set. And it is clearly confirmed in the second histogram of the figure 1. the presence of outliers localized between 5 and 10 as earlier said.
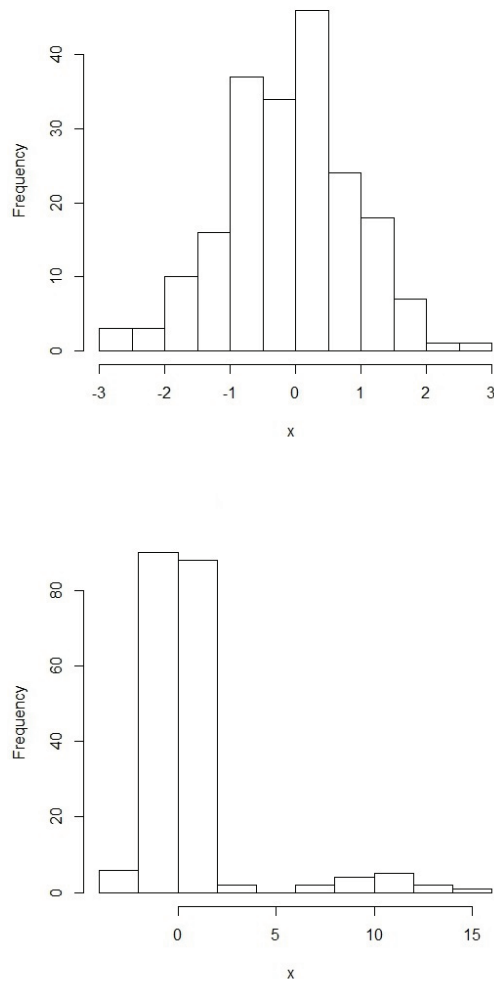
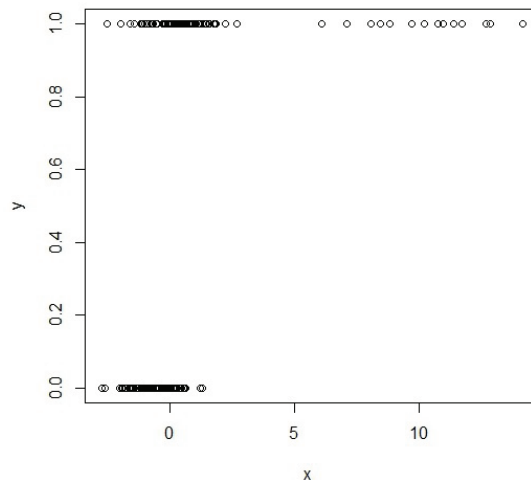Figure 1. Histogram of *X* in both clean and contamination cases

Figure 2. Relationship between Response Variable and Endogenous Variable *X* With outliers

Table 1-2 show the simulated results of all the fitted models for data with various percentages of leverage points. In absence of outliers (0% of lev pt), it can be observed that, there is no significant difference between the restricted logistic and the robust models based on the DIC value. But the SsRLR model seems to give better estimated values of the parameters.

Table 1. Description of variables X and assumed values of parameters manipulated in simulation

Variables and Parameters	Assumed values
n	200
x_1	$x_1 \sim N(n, 0, 1)$
x_2	$x_2 \sim N(n, 0, 1)$
β_0	0
β_1	2
β_2	2
α	$\alpha \sim U(a, b)$ with a and b belonging to [0,1]

The Restricted LR model was immediately affected by 3% of leverage points giving the highest DIC value. Gelman's model was influenced as well showing parameter estimates which were not stronger than the expected one, while the SsRLR model let the data itself to select 2.664E-5 alpha value that improved the parameter estimated values.

It is interesting to observe that the 5% of leverage points do not have effect on the SsRLR model. This latter confirms its robustness giving more better simulated result with the smallest DIC value.

The α values 5.014E-5 and 2.059E-3 respectively self selected in the presence of outliers (5% and 7% of lev pt) allowed the data to minimize the influence of those latter in the parameter estimation.

Based on the criterion that a difference in DIC values from 3, 4 between two models provides a better fit, it can be clearly concluded that the best fitting model is the Self Selecting Robust Logistic Regression (SsRLR) model with small DIC value when there is presence of outliers in the binary response data.

Furthermore, based on the BIC, the SsRLR model with the lowest BIC value is the preferred best fitting model (Schwarz. 1978).

5. Discussion

This study uses Bayesian techniques to develop robust logistic regression model when outliers are present in binary response data. The study develops robust logistic model to help improve parameter estimation fitting. In this study, the approach used in the robust model is based on a trimming value alpha, α chance of random error in both direction of the interval [0, 1].

From the existing contribution of Gelman (2004) that fixed α and $(1 - 2\alpha)$ in his model, we extended by self selecting theses probability values depending on the data at hand and gave them a Uniform $[a, b]$ prior distribution.

In this study, we clearly confirmed that these probability values could also be determined by the data itself. In other words, depending on the binary data at hand, this latter could itself select α and $(1 - 2\alpha)$.

We found that the smaller the assigned values of a and b, the smaller the self selected α and the more efficient the estimates obtained from simulation results will be, compared to the ones otained from both the GRLR and the LR models when the data is either clean or contaminated.

Another finding is that the self selecting robust logistic regression model is better fitting model compared to the Restricted LR model based on DIC value using Bayesian approach implemented in WinBUGS.

The SsRLR model provides a reliable fitting model based on the lowest BIC value compared to the RLR and GRLR models.

We also found that the Restricted LR model has minimized the effect of the outliers present in the data and allowed achievement of better results. Despite this, the Self Selecting Robust Logistic Regression model presented more reliable results in comparison to the Restricted LR contrary to Gelman's robust logistic regression model.

6. Conclusions

This work aims to extend the performance of logistic regression for binary data. Ordinary LR with arbitrary outliers was

Table 2. Simulated results of all models for Data with Leverage Points (0% and 3%)

% of lev pt	0%			3%		
Estimate	SsRLR	GRLR	LR	SsRLR	GRLR	RLR
β_0	0.1074	0.1261	0.1072	0.02201	-0.05214	0.00607
σ_0	0.2237	0.1944	0.223	0.0216	0.2401	0.2034
β_1	2.371	2.201	2.363	2.035	1.714	1.875
σ_1	0.3139	0.3702	0.3618	0.3101	0.2016	0.3185
β_2	2.326	2.107	2.304	2.017	1.556	1.571
σ_2	0.2987	0.3731	0.3704	0.3021	0.263	0.3083
α	1.004E-5	0.01	-	2.664E-5	0.01	-
σ_α	5.835E-6	-	-	2.578E-5	-	-
Dhat	136.124	135.211	139.987	146.573	147.123	152.890
pD	2.917	2.820	2.978	2.765	2.781	3.061
BIC	154.532	153.619	153.793	164.981	165.531	166.696
DIC	142.208	141.958	143.098	153.624	153.686	157.812

Table 3. Simulated results of all models for Data with Leverage Points (5% and 7%)

% of lev pt	5%			7%		
Estimate	SsRLR	GRLR	RLR	SsRLR	GRLR	RLR
β_0	0.06466	0.02474	0.06374	-0.1779	-0.1898	-0.1183
σ_0	0.2315	0.2292	0.2123	0.2202	0.2271	0.2333
β_1	2.194	1.859	2.069	2.344	2.198	2.217
σ_1	0.3014	0.2943	0.4068	0.3097	0.3634	0.4015
β_2	2.289	2.125	2.184	2.25	1.951	2.087
σ_2	0.3722	0.3104	0.3841	0.3475	0.3842	0.3907
α	5.014E-5	0.01	-	0.002059	0.01	-
σ_α	2.888E-5	-	-	0.001145	-	-
Dhat	125.280	126.038	131.056	113.339	114.058	119.009
pD	2.849	2.898	2.845	2.801	2.943	3.077
BIC	143.688	144.446	144.862	131.747	132.466	132.815
DIC	132.199	132.376	136.645	119.395	119.524	124.164

shown to fail. We proposed a robust SsRLR model that dealt with such contamination. It was also observed that by fixing the value of alpha, GRLR model was not that robust to the influential observations.

We proposed in that study a novel robust (LR) model to solve this issue. To proceed, we developed a self selecting robust logistic model, then investigated the robustness of this latter. We proposed a clear way of specifying the trimming values as required by the user, as opposed to fixing it.

One finding indicated across the simulation results that SsRLR model performs well in its specificity of letting the binary data itself to select the alpha value necessary to better improve the quality of the parameter estimates. Based on the smallest DIC and BIC value respectively, our SsRLR model was found the best fitting model under contaminated binary data sets.

We found that as long as α value is smallest self selected by the data at hand, the robustness of the SsRLR model is more improved. That is our contribution to Gelman's robust logistic regression model.

7. Recommendation

By carrying out this study, we showed the insufficiency contained in the Gelman's model when it comes to dealing with binary response data in presence of leverage points.

From there, we recommend the statistician to use the SsRLR model when modeling binary data in the case of covariate corruption and to further investigate its robustness in future research.

We also recommend the future researchers to focus more on the robustness of the GRLR model by studying the behavior of the trimmed probability alpha when the outliers occur in the Y-space called misclassification-type error or in both Y-space and X-space. The next author can widen this work by discussing the problem of improper prior of α.

Acknowledgments

References

Copas, J. B. (1988). Binary regression models for contaminated data (With discussion): *Journal of the Royal Statistical Society. Series B (Methodological)*, 2(50), 225-265.

Croux, Christophe Haesbroeck, Gentiane. (2003). *Implementing the Bianco and Yohai estimator for logistic regression: Computational Statistics and Data Analysis, 1-2*(44), 273-295. https://doi.org/10.1016/S0167-9473(03)00042-2

Edgeworth, F. Y. (1887). On Observations Relating to Several Quantities: Hermathena, 279-285.

Gelman, Andrew. (2004) *Detection of Outliers and Influential Observations in Binary Logistic Regression: Journal of the American Statistical Association, 99*(466), 537-545.

Geman, S., & Geman, D. (1984) *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images: IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6)(PAMI-6), 721-741.

Huber, P. J. (1973). *Robust Regression: Asymptotics, Conjectures and Monte Carlo: Annu. Rev. Ecol. Syst.*, 1(5), 799-821. https://doi.org/10.1214/aos/1176342503

Huelsenbeck, J. P., & Crandall, K. A. (1997). *Phylogeny Estimation and Hypothesis Testing Using Maximum Likelihood: Annu. Rev. Ecol. Syst.*, 1(28), 437-466. https://doi.org/10.1146/annurev.ecolsys.28.1.437

Koenker, R., & Bassett, G. (1985). *On Boscovich's Estimator: Ann. Stat.*, 4(13), 1625-1628. https://doi.org/10.1214/aos/1176349759

Pregibon, D. (1981) *Logistic Regression Diagnostics: The Annals of Statistics*, 9(4), 705-724. https://doi.org/10.1214/aos/1176345513

Robert, Christian Ntzoufras, Ioannis. (2012). *Bayesian Modeling Using WinBUGS: Chance*, (2)(25), 60-61.

Rousseeuw, P. J., & Christmann, A. (2003). *Robustness against separation and outliers in logistic regression: Computational Statistics And Data Analysis*, 3(43), 315-332. https://doi.org/10.1016/S0167-9473(02)00304-3

Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection: J. Wiley* https://doi.org/10.1002/0471725382

Rousseeuw, P. J. (1984). *Least Median of Squares Regression: Journal of the American Statistical Association*, 79(388), 871-880. https://doi.org/10.1080/01621459.1984.10477105

Rousseeuw, P. J., & van Driessen, K. (1999). *Least Median of Squares Regression: Technometrics*, 3(41), 212. https://doi.org/10.1080/00401706.1999.10485670

Sarkar, S. K., Midi, H., & Rana, S. (2011). *Parameterization and Bayesian Modeling: The Annals of Statistics*, 9(4), 705-724.

Schwarz, Gideon. (1978). *Estimating the Dimension of a Model: The Annals of Statistics*, 2(6), 461-464. https://doi.org/10.1214/aos/1176344136

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, Angelika. (2002).*Bayesian measures of model complexity and fit: Journal of the Royal Statistical Society Series B (Statistical Methodology, 4*(64), 583-639. https://doi.org/10.1111/1467-9868.00353

Tukey, J. W. (1960). *A survey of sampling from contaminated distributions: Stanford University Press, Contributions to Probability and Statistics*, 448-485.

Victoria-Feser, Maria-Pia. (2002). *Robust inference with binary data: Psychometrika*, *1*(67), 21-32. https://doi.org/10.1007/BF02294707

Yohai, V. J. (1987). *High Breakdown-Point And High Efficiency Robust Estimates For Regression: The Annals of Statistics*, *2*(15), 642-656. https://doi.org/10.1214/aos/1176350366

Appendix

Softwares Used in the Study.

R software, WinBUGS.

Copyrights