Panamerican Mathematical Journal
ISSN: 1064-9735
Vol 35 No. 3s (2025)

Robust Regression Analysis for Pregnant Women Nutrition Level of the State Tamil Nadu

D. Karunanidhi[1,*] and S. Sasikala[2]

[1,*]Research Scholar, Department of Statistics, Thanthai Periyar Government Arts & Science College (Autonomous) (Affiliated to Bharathidasan University), Trichy -23, Tamil Nadu, India. Email: karunamartin78@gmail.com

[2]Assistant Professor, Department of Statistics, Thanthai Periyar Government Arts & Science College (Autonomous) (Affiliated to Bharathidasan University), Trichy -23, Tamil Nadu, India. Email: harisasi24@gmail.com

Article History:

Received: *04-10-2024*

Revised: *26-11-2024*

Accepted: *09-12-2024*

Abstract:
Robust regression analysis is an analysis which is used if the outlier present in the regression model. Outlier causes the data to be abnormal. The ordinary least square method is commonly used for estimating the parameter. The estimated model with outliers is a biased model, which resulted with the unreliable result. For this, the robust regression model is an effective model when the outliers presence in the dataset. In this study, Robust M-estimation can estimate the parameter model and the nutrition conditions of pregnant women Tamil Nadu 2011-2019 has been considered. This study aims to study the OLS method and Robust M-estimation method on data with 5% and 10% level of significance and the outliers are present in the dataset, the robust method resulted that the robust regression method is the best method to handle the data set.

Keywords: Nutrition data, M-estimation, Ordinary Least square method, R^2 values

1. Introduction

Regression analysis is a statistical analysis that is used to determine the relationship between response variables and predictor variables. According to the definition, “Regression analysis is used to determine conclusions from data that has related relationship between response and predictor variables”. The data under study is having more than one predictor variables; the multiple regression analysis has been used. The robust method is used to study the data which is having outliers in it and also used when the data not follows the normality which affects the parameter estimation. The main objective of the robust method is to detect outliers in the data and results with reliability. In robust estimation, there are number of estimators available, namely, RM-estimator, LMS-estimator, LTS-estimator, S-estimator, LWS-estimator, M-estimator, LAD estimator, MM estimator and REWLSE estimator. In this study, the analysis robust LAD estimator and M-estimator are used. Those estimators are simplest estimators of robust estimation.

Outliers of the given dataset must be estimated before estimate the robust model. There are number of ways available to detect the outliers in given data set such as, Stundetized Deleted Residual (TRES) and Cook distance methods. In this study, Cook distance measure has been used at 5% and 10% level of significance.

This study will examine how the robust M-estimator and OLS method are compared at 5% and 10% level of significance. Here, there are 3 predictor variables in the given dataset such as, age-wise

nutrition level of pregnant women, number of health centres and pregnant women population of the state.

2. Review of Literature

There are number of literatures are available and number of authors studied about the robust methods. The following studies are taken into account for the present study.

Ko D.J and Chang. T [1] discussed the influence functions and asymptotic distributions of such estimators and studied the necessary and sufficient conditions under which are optimal in a sense to Hampel method. The proof for the unbiasedness of L-norm and M-estimators are given for regression models.

Seija Sirkia, Sara Taskinen and Hannnu Oja [2] analyzed a family of symmetrised M-estimators of multivariate scatter and also discussed the multivariate elliptical case to consider the robustness and efficiency properties of estimators.

Kris Boudt and Christophe Croux [3] studied the use of volatility models with the property of bounded innovation propagation using robust M-estimators. The influence function approach to robust estimation of parametric models and resulted for M-estimators with a bounded influence function is generalized to allow for arbitrary choices of the asymptotic efficiency criterion and the norm of the influence function.

Krzysztof Czaplewski and Mariusz Waz [4] studied the M-estimation methods that are used in geodesy as well as simple neural networks and proposed automating the process of determining a vessel's position at sea by using comparative navigation methods.

Beatriz Sinova, Gil Gonzalez Rodriguez and Stefan Van Aeist [5] studied the properties of M-estimators and also studied the consistency of the estimators and robustness by the means of their breakdown point and their influence function. The M-estimators are also empirically compared to trimmed means for functional data [6, 7].

The numerical computations of the estimates, the determination of their tuning parameters, and the evaluation of measures that characterize their distribution. A detailed asymptotic theory of estimating a location parameter for contaminated normal distribution, and exhibits estimators intermediaries between sample mean and sample median that are asymptotically most robust among all translation invariant estimators.

3. Research Methodology

This study used secondary data which are available in the Tamil Nadu health department website. This data consists of number of variables regarding the pregnant women. Here for this study, we are considered the age-wise nutrition level of pregnant women, number of health centre's and pregnant women population of the state for the period of 2011-2019. The analysis classified into following categories:

1. Testing of normality of the given dataset
2. Testing the spatial heterogeneity

3. Outlier detection
4. Estimating the parameter of the model
5. Testing estimators of the model
6. Testing for residual normality of the model

4. Result and Discussion

4.1 Normality test

The Kolmogrov Smirnov test has been considered to test the normality of the given dataset.

Hypothesis setting: H_0: the classical linear models are normally distributed.

H_1: the classical linear models are not normally distributed.

The test statistic value T_n determined as follows:

Table 1. Normality test results of Nutrition level of pregnant women

Year	T_n	P-value
2011	0.6156	2.12e-10
2012	0.6215	2.44 e-9
2013	0.7107	2.78 e-9
2014	0.5508	1.78 e-5
2015	0.5678	1.40 e-5
2016	0.5678	1.40 e-5
2017	0.6743	2.56 e-9
2018	0.8705	3.23 e-9
2019	0.7680	2.92 e-9

Hence, the p-value of the test statistic is less than 0.05. Therefore, H_0 rejected. It is concluded that, the given data residuals are not normally distributed due to the existence of outliers.

4.2 Robust Regression

Outlier detection is done using Cook distance method. Then the results are compared for 5% and 10% level of significance.

Table 2. Results of detecting Outliers with Cook distance method

Year	Alpha 5%	Alpha 10%
2011	3(11,10,20)	2 (11,20)
2012	2(11,12)	4(10,8,12,11)

2013	4(12,11,9,8)	3(12,11,9)
2014	2(18,11)	3 (11,10,18)
2015	2(19,20)	4(21,19,12,10)
2016	2(13,18)	3(11,13,18)
2017	2(21,17)	2(21,17)
2018	3(23,12,10)	1(23)
2019	2(12,11)	3(12,11,9)

Table 2 revealed that the nutrition level of pregnant women TamilNadu from 2011-2019 has outliers. Here in this study 5% level of significance resulted with the best results.

By using R software, Table 3 compared the results which attained by OLS and Robust M-estimator and it is resulted that the R^2 values of robust M-estimator. The R^2 value of robust M-estimator having higher than the OLS R^2 value. It is concluded that the robust M-estimator explain the response variable better than OLS estimator when data contains outliers.

Table 3. R^2 value results in outlier data at 5% level of significance

Year	R^2 OLS	R^2 Robust M-estimator
2011	0.1458	0.2058
2012	0.1643	0.2156
2013	0.1756	0.1987
2014	0.1876	0.2174
2015	0.2138	0.2564
2016	0.2223	0.2786
2017	0.2148	0.2340
2018	0.2148	0.2340
2019	0.2250	0.2458

Table 4. Robust M- estimation of the data set at 5% level of significance

Year	Maximum Iteration	β_0	β_1	β_2	β_3
2011	11	-33.565	4.1567	0.0051	0.045
2012	11	-30.450	5.1108	0.0035	0.051
2013	11	-28.450	4.4280	0.040	0.010
2014	12	-41.330	4.3208	0.075	0.025
2015	13	-21.340	4.6508	0.005	0.042
2016	13	-23.650	4.7575	0.008	0.032
2017	11	-42.220	4.6590	0.0045	0.025
2018	11	-21.650	3.7163	0.0032	0.015

2019	11	-28.110	4.3208	0.006	0.010

Table 5. OLS estimation of the data set at 5% level of significance

Year	β_0	β_1	β_2	β_3
2011	-28.356	4.1140	0.0044	0.015
2012	-23.569	5.1008	0.0031	0.011
2013	-22.340	4.3250	0.001	0.050
2014	-32.480	4.1280	0.068	0.035
2015	-17.890	4.1200	0.0043	0.012
2016	-21.780	4.0082	0.006	0.042
2017	-35.467	4.2305	0.0020	0.065
2018	-21.450	3.1105	0.0015	0.095
2019	-27.890	3.9803	0.0030	0.020

From Table 4, we have obtained the parameter estimation difference is less than $1*10^{-5}$ at the end of the iteration and it is concluded that robust M-estimation method attained reliable results in the outlier data set at 5% level of significance.

This study revealed that there was an influence between the response variable and overall predictor variable with significant results with the table value 3.432. Table 6 describes the values of F values of the model parameters with 5% level of significance.

Table 6. Robust M- estimation of the data set at 5% level of significance

Year	β_0	β_1	β_2	β_3	F value
2011	-33.565	4.1567	0.0051	0.045	4.5430
2012	-30.450	5.1108	0.0035	0.051	5.0805
2013	-28.450	4.4280	0.040	0.010	4.7854
2014	-41.330	4.3208	0.075	0.025	5.1100
2015	-21.340	4.6508	0.005	0.042	4.5504
2016	-23.650	4.7575	0.008	0.032	5.4103
2017	-42.220	4.6590	0.0045	0.025	4.3709
2018	-21.650	3.7163	0.0032	0.015	3.4560
2019	-28.110	4.3208	0.006	0.010	4.1145

5. Conclusions

The regression models are used to evaluate the model based on the predictor and response variable. Generally, the simplest methods OLS (ordinary Least Square) method is used. In most of the cases, it may leads biased results. For example, outliers presence data, non-normality data. In that type of cases robust estimated models help us to attained the reliable results. In this study, the considered data has the outliers and the non-normality. The both estimation methods are studied and resulted that

the model with parameter estimators obtained from M-estimator method can be effectively used to predict the nutrition level of pregnant women. The R^2 value of M-estimator model produces the significant values rather than R^2 value of OLS estimator. The study resulted that the best model for outlier dataset is robust M-estimator and it is helpful for us to attain the best results for the given dataset.

References

[1] Ko, D. J., & Chang, T. (1993). Robust estimators on spheres. *Journal of Multivariate Analysis*, *45*(1), 104-136.
[2] Sirkia, S., Taskinen, S., & Oja, H. (2007). Symmetrized M-estimators of multivariate scatter. *Journal of Multivariate Analysis.*
[3] Boudt, K., & Croux, C. (2010). Robust M-estimation of multivariate GARCH models. *Computational Statistics and Data Analysis*, *54*(11), 2459-2469.
[4] Czaplewski, K., & Waz, M. (2017). Improvement in accuracy of determining a vessel's position with the use of neural networks and robust M-estimation. *Polish Maritime Research*, *24*(1), 93.
[5] Sinova, B., Gonzalez Rodriguez, G., & Van Aeist, S. (2018). M-estimators of location for functional data. *Bernoulli Society for Mathematical Statistics and Probability*, *24*(3), 2328-2357.
[6] Machado, J. A. F. (1993). Robust model selection and M-estimation. *Econometric Theory*, *9*, 478-493.
[7] Susanti, Y., Pratiwi, H., Sulistijowati, S., & Liana, T. (2014). M-estimation, S-estimation, and MM-estimation in robust regression. *International Journal of Pure and Applied Mathematics*, *91*(3), 349-360.