Research Journal of Mathematical and Statistical Sciences ISSN 2320-6047
Vol. **7(1),** 1-6, January **(2019)** *Res. J. Mathematical and Statistical Sci.*

Statistical diagnostics of models for binomial response

Ali Choo, Yu Feng*, Zamir Ali and Munsir Ali
School of Science, Department of Probability and mathematical Statistics Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, P.R., China
alistat94@njust.edu.cn

Available online at: www.isca.in, www.isca.me
Received 22nd October 2018, revised 26th December 2018, accepted 5th January 2019

Abstract

In regression analysis, outliers are frequently encountered. Diagnostics of outliers is an essential tool of the model building process. Most of the time analysts depend on ordinary least square (OLS) method without identifying outliers. It is evident that OLS utterly fails in the identification of outliers. In this paper, we use diagnostics techniques to detect residuals and influential points in statistical models for binomial the response. Gauss-Newton and likelihood distance methods were considered to identify the outliers in parameter estimation in non-linear regression analysis. The results illustrated single and multiple outliers in dataset.

Keywords: Residuals, logistic regression, log-likelihood distance, hat matrix, non-linear regression models, 2010 Mathematics subject classification (62J99, 62J02, 62F99, 62J20, 62G05).

Introduction

Binomial response is a method in which the dependent variable Y is the outcome of a sequence of a binomial trail, on the other hand a line of one of two consistent disjoint results (expressed as 1 for "success" and 0 for "failure"). In the binomial response, the possibility of success is linked to the independent variables (explanatory variables). Binary regression molds are equivalent as the discrete choice models. There has been vast advancement of diagnostics degrees for binomial the response model adapted by (ML) maximum likelihood. Whilst there is a testimony that likelihood inference, minimum chi-square estimates, case deletion, likelihood distance and Hellingger distance estimates have beneficial asymptotic and tiny random sample characteristics in a bit frames, and there was no printed trail to enlarge diagnostic techniques placed on these, or in other estimation techniques[1].

In general, the Binomial response models are usually delineated in points of GLM (generalized linear model), a trail to generalize different linear regression model[2,3]. In the binary regression model, nevertheless, apportion with the latent and error inconstant mutable and suppose that the selection itself is a haphazard variable with a linked function. Moreover, this process converts the normal value of the haphazard variable into a value, which is prophesied by the linear predictor. There has been introduced that the two are similar, in the situation of binary superior models: i. Link Function compared to the quintile function of distribution of the error inconstant, ii. inverse linked function to the CDF (cumulative distribution function) of the error inconstant. Suppose that first case of ungrouped binary response is denoted by [0, 1]. When the relationship among x and y is unknown the parameters is linear, the framework is that of the classical linear mixed model. In this case, Bayesian inferential technique is providing satisfactory hierarchical linear model[4,5]. There have been some backgrounds studies about the hierarchical linear model[6]. In practice, the outliers may additionally occur for various reasons including gadget mistakes, the discrepancies in statistics transcription and a few different cases[7]. The idea has been, adopted in light of the previous studies to expand some strategies for the detection of outliers' indifference among situations[8]. For instance, the idea is to rate the influence of outlier, which is based entirely on the difference between two log-likelihood feature with their maximum likelihood estimates the use of the whole facts set with ability outlier (CHO and Tse)[7-9]. The results of the test statistics may be asymptotical as compared to a χ^2 distribution (Cook and Weisberg)[10-17]. Diagnostics techniques needs iterative assessment of MLE, and one-step approximations that are normally enough to understand the foremost critical instances.

In this study, we aim to signify residuals facts points in models for binomial response data and parameters estimation. Further, we present a graphical and diagnostics approach that determine the impact of the estimation technique on the parameter estimates. We have used Gauss-Newton and log-likelihood distance technique with few useful examples for parameters estimation and single and multiple outliers' detection. The conclusion is presented at last.

The models and the parameters estimation

The generalized linear models with the regression analysis are based on likelihoods. We have described the fundamental inferential appliances for parameter estimation the goodness of fit tests and hypothesis testing, since more specified material on model checking and model choice. The likelihood technique interest's grounds of statistical inference and they are often entreated in arguments about exact statistical reasoning.

Let us suppose that $f(x|\theta)$ is conditional distribution for X given the unknown parameter θ. For observed data, $X = x$, the function $\ell(\theta) = f(x|\theta)$, is determined as a function $\theta$and it is known as likelihood function.

Given that the random sample $y_1, \ldots, y_i, \ldots$, mutually with the covariates $x_1, \ldots, x_i, \ldots$, on the other hand the design vector $z_1, \ldots, z_i, \ldots$, is maximum likelihood estimator of the unascertained parameters vectors β in this model $E(y_i|x_i) = \mu_i = h(\acute{z}\,\beta)$ is obtained, by maximizing the likelihood estimation[16]. To deal the cases, of particular data $(i = 1 \ldots n)$ and the grouped information $(i = 1 \ldots, g)$ concurrently. We will exclude $n\; or\; g$ as the maximum limit summation indications. Consequently, sum total may series over i from 1 to n or 1 to g and the weights ωi, have to be the set similar to 1 for the particular data and similar to n_i for the grouped data.

Assumed that the parameter scale ϕ, is already given. In as, much as ϕ emerges as a part in the likelihood estimation, it will be possible to set $\phi = 1$, in this situation without damage of the generality assuming that we are only curious in a point estimate, of β. Only the mean, $\mu_i = h(\acute{z}_i\beta)$ and the variance function, are then accurately determined, so that one has to begin with the appearance for the score function, $s(\beta)$ as an alternative of the log-likelihood $l(\beta)$.

To prevent supplementary complexities representing parameters identifiability, it is presumed at this time on that the "splendid" design matrix.

$Z = (z_1, \ldots, z_i, \ldots, \acute{)}$ Has (full) rank p,

On the other hand, equivalently,

$\sum_i z_i \acute{z} = \acute{Z}Z$ Has rank p.

For binomial scaled responses, i.e., the relative frequency $\bar{y}i$ in addition to repetition number n_i here we have

$$li(\pi_i) = n_i(\bar{y}_i \log\pi_i + (1 - \bar{y}_i)\log(1 - \pi_i))$$

So, inserting the mean composition $\mu_i = h(z_i'\beta)$ eventually provides

$$l_i(\beta) = l_i(h(z_i'\beta)$$

According to the log it model

$$l(\beta) = \sum_i l_i(h(z_i'\beta) \qquad (1)$$

The score function indicates how sensitive a likelihood function $(l(\theta; X)$ is to its parameter $\theta.s$

$$(\beta) = \sum_i \frac{\partial l_i(\beta)}{\partial \beta} \qquad (2)$$

The equation (1) shows the log-likelihood function and the equation (2) demonstrates the score function with response on the ith individuals.

In case of nonlinear regression $\varepsilon_i \sim N(0, \Sigma_i)$ then the y on the parmeter β of the observation information matrix $-\ddot{L}(\beta)$, score function $\dot{L}(\beta)$ and the fisher information matrix $I(\beta)$ respectively. The other iterative approaches have applied, to solve the likelihood equation[1]. Moreover, the Newton-Raphson[18] approach obtained from Fisher scoring, if the expected information $F(\beta)$ changed by observed, information $F_{obs}(\beta)$. Computational, of nonlinear least square estimate the need to apply the iterative numerical set of rules.

$\dot{L}(\hat{\beta}) = 0$, We may additionally use the Taylor[19] expansion at the point β_0.

$$\beta^{i+1} = L(\beta^i) + [-\ddot{L}(\beta^i)]^{-1}\dot{L}(\beta^i), \quad i = 1, 2, \ldots \qquad (3)$$

Until $|\beta^{i+1} - \beta^i| < \delta$, δ is an advance constant value. β^{i+1} will converge to $\hat{\beta}$ under some regular condition and, the pace of convergence will depend on the choose value of β_o.

Statistical diagnostics for models for binomial responses

Diagnostics techniques for assessing, the fit of classical linear regression models are commonly or ordinarily used[2-11]. Almost, all these tools based on hat matrix, case deletion measure, graphical, and a presentations of residuals[12].

A general approach to influence: The methodology, of local influence has a few circumstances[13]. It is pulling in, as it take into account for computing the influence of a solitary observation, additionally evaluation of the impact of numerous observations. Persuasive observations are firmly associated with high use perceptions/observations and outliers[14]. A more profound comprehension of the diagnostic measures is utilized to distinguish influential observations can be accomplished when they are dissected concerning high-use observations, and outliers[11].

The influence measures of the $i - th$ case on the ML estimate $\hat{\beta}$ can be based on the sample.

Influence cure $SIC_i \propto \hat{\beta} - \hat{\beta}_{(i)}, \widehat{\beta}_{(i)}$ represented the ML estimate of β estimated without the $i - th$ case. At the same time that this thought is clear, this might be computationally extravagant to, carry out since $n + 1ML$ gauges are required. This might be valuable to think about a quadratic approximation of; $L_{(i)}$ the probability got subsequent to the evacuating the $i - th$ case.

$$L_{(i)} \cong L_{(i)}(\hat{\beta}) + (\beta - \hat{\beta})^T \dot{L}_{(i)}(\hat{\beta}) + \frac{1}{2}(\beta - \hat{\beta})^T \ddot{L}_{(i)}(\hat{\beta})(\beta - \hat{\beta}) \qquad (4)$$

Where: $L_{(i)}(\hat{\beta})$ is the gradient vector with $j-th$ element $\partial L_{(i)}(\beta)/\partial\beta_j$ estimated $\beta = \hat{\theta}$.

If $-L_{(i)}(\hat{\beta})$ is positive definite the quadratic approximation is maximized at,

$$\hat{\beta}^1_{(i)} = \hat{\beta} - (\ddot{L}_{(i)}(\hat{\beta}))^{-1}\dot{L}_{(i)}(\hat{\beta}) \quad (5)$$

Log-likelihood distance LD_i as

$$LD_i = 2\left[L(\hat{\beta}) - L(\hat{\beta}_{(i)})\right] \quad (6)$$

On other hand, using the one-step estimator,

$$LD^1_{(i)} = 2\left[L(\hat{\beta}) - L(\hat{\beta}^1_{(i)})\right] \quad (7)$$

It has easily observed to be the general class with $t(\beta) = L(\beta)$, and $LD_{(i)}$ is not essentially a component of simply the sample impact bend for β. The proportions of LD_i and LD^1_i may likewise be illuminated in the terms of the asymptotic certainty region.

$$\{\beta: 2\left[L(\hat{\beta}) - L(\beta)\right] \le \chi^2(\alpha; q)\}$$

Where, $\chi^2(\alpha; q)$ is the upper α purpose of the chi-square distribution, with $q\, df, and\, q$ is the dimension of β. Log-likelihood distance can accordingly, be aligned by correlation with the $\chi^2\,(q)$ distributions.

If the log-likelihood shapes are roughly, elliptical then LD_i can be conveniently approximated by Taylor expansion of $L(\hat{\beta}_{(i)})$ around $\hat{\beta}$,

$$L(\hat{\beta}_{(i)} \cong L(\hat{\beta}) + (\beta_{(i)} - \hat{\beta})^T\dot{L}(\hat{\beta}) + \frac{1}{2}(\hat{\beta}_{(i)} - \hat{\beta})^T\left(\ddot{L}(\hat{\beta})\right)(\hat{\beta}_{(i)} - \hat{\beta})$$

And, seeing that $\dot{L}(\hat{\beta}) = 0$,

$$LD_i \cong (\hat{\beta}_{(i)} - \hat{\beta})^T\left(-\ddot{L}(\hat{\beta})\right)(\hat{\beta}_{(i)} - \hat{\beta}) \quad (8)$$

A particular approximation can be acquired by supplanting the observed data, $-\ddot{L}(\hat{\beta})$ in equation (4) by the expected information matrix, estimated at $\hat{\beta}$.

Logistic regression and generalized linear models: The output variable y only takes on value $\in (0,1)$ for binary classification, we need a different way of representing $E\left(\frac{y}{x}\right)$ so that the range of $y \in (0,1)$. One convenient form to use is the sigmoid or logistic function[15]. Let us suppose a vector valued input variable $x = (x_1, .., x_p)$. The logistic function is S formed and approaches $0 (as\ x \to -\infty)\ or\ 1\ (as\ x \to \infty)$.

Suppose a sample $Y^T = (y_1, y_2, \dots, y_n)$ of free, irregular factors to such an extent that y_i is binomially disseminated $B(n_j, p_j)$ with n_j known and p_j obscure. The logistic regression, model indicates the relationship,

$$n_j = \text{logit}(p_j) = \log\left[\frac{p_j}{1-p_j}\right] = X_j^T\beta, \quad j = 1,2,\dots,n \quad (9)$$

Where, $X_1, X_2, \dots, X_n$ are $p' - vectors$ of illustrative, factors and β is an obscure parameter vector. In such model's estimation of β is ordinarily, a noteworthy concern.

The log-probability for $\eta = X\beta$ is,

$$L(\eta) = L(X\beta) = \sum_{j=1}^{n}[y_i X_j^T\beta - a_j(X_j^T\beta) + b_j(y_j)] \quad (10)$$

Where: $a_j(z) = n_j \log[1 + \exp(z)]$ and $b_j(z) = \log\binom{n_j}{z}$. The maximum likelihood estimates $\hat{\beta}$ of β is usually found applying Newton's method. Using different notation defines $\hat{p}_j = \exp(X_j^T\hat{\beta})/[1 + \exp(X_j^T\hat{\beta})]$ and let $\widehat{W}$ be a $n \times n$ diagonal matrix with $j-th$ diagonal $n_j\hat{p}_j(1-\hat{p}_j)$. Also suppose $\hat{s}$ be a $n - vector$ with $j-th$ element $\hat{s}_j = y_j - n_j\hat{p}_j$. One can show that,

$$\dot{L}_{(i)}(\hat{\eta}) = X^T_{(i)}\hat{s}_{(i)}; \quad \ddot{L}_{(i)}(\hat{\eta}) = -(X^T_{(i)}\widehat{W}_{(i)}X_{(i)}) \quad (11)$$

By using (10), finally we get this equation,

$$\hat{\beta}^1_{(i)} = \hat{\beta} - \frac{(X^T\widehat{W}X)^{-1}x_i\hat{s}_i}{1-\hat{v}_{ii}} \quad (12)$$

Where: $\hat{v}_{ii}$ is the $i-th$ diagonal element of $\hat{V} = \widehat{W}^{\frac{1}{2}}X(X^T\widehat{W}X)^{-1}X^T\widehat{W}^{\frac{1}{2}}$. Pregibon[20] discusses, the accuracy of this one-advance guess and presumes that component wise, the estimation will in general think little of completely iterated esteem yet this might be insignificant for recognizing, compelling cases.

Measures, for the distinctions $\hat{\beta} - \hat{\beta}_{(i)}\ or\ \hat{\beta} - \hat{\beta}_{(i)}$ can be determined applying circular approximation probability removal or alteration in fitted esteem vectors as talked about beneath. Following Pregibon[20] we will allow for these to characterize, an influence for

$$LD^I_{(ij)}(\beta) = (\hat{\beta} - \hat{\beta}^I_{(ij)})^T\,(-\ddot{L}(\hat{\beta}) . (\hat{\beta} - \hat{\beta}^I_{(ij)}) \quad (13)$$

We give extension for log likelihood distance LD_i for binomial response data.

One case of likelihood distance

$$LD_{ij} = 2[L(\hat{\beta}) - (\hat{\beta}_{(ij)})] \quad (14)$$

$$\mathrm{LD}^{\mathrm{I}}_{(\mathrm{ij})}(\beta) = (\hat{\beta} - \hat{\beta}^{\mathrm{I}}_{(\mathrm{ij})})^{\mathrm{T}}\,(-\ddot{\mathrm{L}}(\hat{\beta}) . (\hat{\beta} - \hat{\beta}^{\mathrm{I}}_{(\mathrm{ij})})$$

Putting the value of $\hat{\beta} - \hat{\beta}^1_{(i)}$ in equation (12) leads to the form,

$$LD^{I}_{(ij)}(\beta) = \left(\frac{\left[-\frac{\partial s(\beta)}{\partial \beta'}\right]^{-1} \sum_{ij} y_{ij} \left(1 - \frac{e^{z_i'\beta}}{1+e^{z_i'\beta}}\right) z_{ij}'}{1-\hat{v}_{ii}} \right)^{T} - \frac{\partial s(\beta)}{\partial \beta'} \left(\frac{\left[-\frac{\partial s(\beta)}{\partial \beta'}\right]^{-1} \sum_{ij} y_{ij} \left(1 - \frac{e^{z_i'\beta}}{1+e^{z_i'\beta}}\right) z_{ij}'}{1-\hat{v}_{ii}} \right) \quad (15)$$

Multiple case of likelihood distance,

$$LD_i = 2[L(\hat{\beta}) - (\hat{\beta}_{(i)})] \quad (16)$$

$$LD^{I}_{(i)}(\beta) = (\hat{\beta} - \hat{\beta}^{I}_{(i)})^{T} (-\ddot{L}(\hat{\beta}) . (\hat{\beta} - \hat{\beta}^{I}_{(i)})$$

Finally, substituting into (12), this form becomes

$$LD^{I}_{(i)}(\beta) = \left(\frac{\left[-\frac{\partial s(\beta)}{\partial \beta'}\right]^{-1} \sum_{ij} y_{i} \left(1 - \frac{e^{z_i'\beta}}{1+e^{z_i'\beta}}\right) z_{i}'}{1-\hat{v}_{ii}} \right)^{T} - \frac{\partial s(\beta)}{\partial \beta'} \left(\frac{\left[-\frac{\partial s(\beta)}{\partial \beta'}\right]^{-1} \sum_{ij} y_{i} \left(1 - \frac{e^{z_i'\beta}}{1+e^{z_i'\beta}}\right) z_{i}'}{1-\hat{v}_{ii}} \right) \quad (17)$$

Example-1: We examine the information in the below Table-1 that taken from a study reported by Weisberg (1980α)[17]. In an examination to explore the measure of a medication held in the liver of rodent, 19 rodents were arbitrarily chosen, gauged, put under light ether anesthesia and given an oral dose of a medication. The portion a creature got was resolved as around 40mg of the medication per kilogram of body weight, since liver weight is known which is emphatically identified with the body weight and it was felt that huge livers would assimilate to a greater, extent a given dose than littler organs. After a settled length of the time, each rodent was yielded, the liver weight, and the level of the dose in the liver was resolved.

Table-1: Rat data Source: Wesiberg (1990α)[17].

Y	X_1 Body weight (g)	X_2 Liver weight (g)	X_3 Relative dose
0.42	176	6.5	0.88
0.25	176	9.5	0.88
0.56	190	9.0	1.00
0.23	176	8.9	0.88
0.23	200	7.2	1.00
0.32	167	8.9	0.83
0.37	188	8.0	0.93
0.41	195	10.0	0.98
0.33	176	8.00	0.88
0.38	165	7.9	0.84
0.27	158	6.9	0.80
0.36	148	7.3	0.74
0.21	149	5.2	0.75
0.28	163	8.4	0.81
0.34	170	7.2	0.85
0.28	186	6.8	0.94
0.30	146	7.3	0.73
0.37	181	9.0	0.90
0.46	149	6.4	0.75

Herein, we consider a bio-exponential characteristic to calculate Gauss Newton method,

$$y = \beta_1 \exp(-\beta_2 x) + \beta_3 \exp(-\beta_4 x), \beta_1, \ldots. \beta_4 > 0, \quad (18)$$

We observe Gauss Newton method,

$$\hat{\beta}^{(k+1)} = \sum_i y_i (log e^{z_i'\beta^{(k)}} - \log(1 + e^{z_i'\beta^{(k)}})) + (1 - y_i) + \left[-\frac{\partial s(\beta)}{\partial \beta'}\right]^{-1} \sum_i y_i \left(1 - \frac{e^{z_i'\beta^{(k)}}}{1+e^{z_i'\beta^{(k)}}}\right) z_i' \quad (19)$$

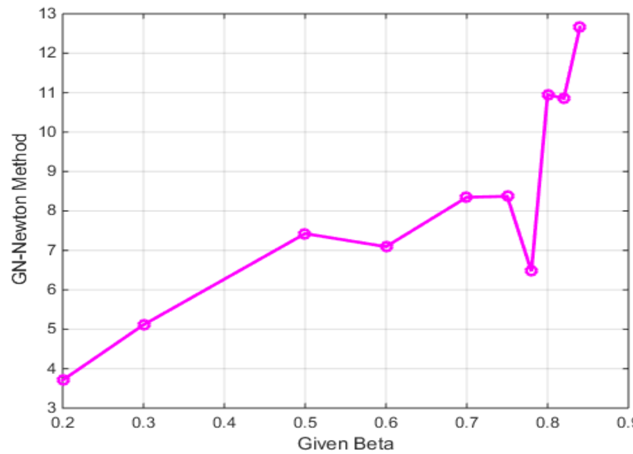

Figure-1a: Plot for the table-1 under the model (19).

We are using MATLAB, to solve this problem. Here, we chose the initial values of $\beta, \beta_o = [0.11, 0.20, 0.40, 0.50]$, after five iteration we get $\hat{\beta} = [3.70, 5.10, 7.40, 7.09]$. Which can be satisfied under property $\|\beta^{i+1} - \beta^{i}\| < 10^{-4}$ biexponential regression feature to compute Gauss Newton. The result of this estimation of the parameter is primarily based on 19 responses for the third subject, which has been given in the above table.

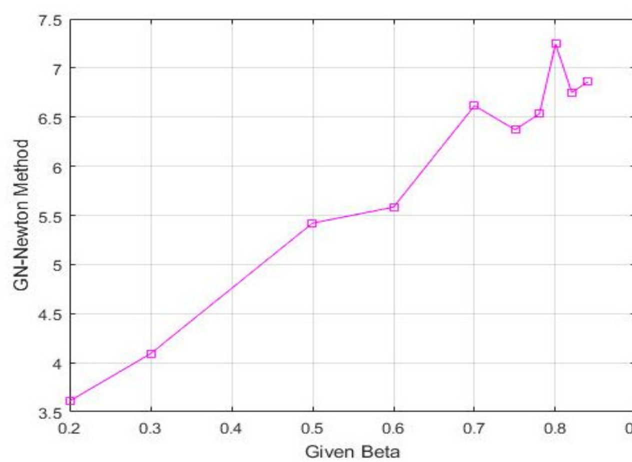

Figure-1b: Plot for the table 1 under the model (19).

We chose the initial values of $\beta, \beta_o = [0.200, 0.300, 0.500, 0.600]$, after five iteration we get $\hat{\beta} = [1.260, 1.180, 1.20, 1.150]$. Which is satisfied under circumstance $\|\beta^{i+1} - \beta^{i}\| < 10^{-4}$.

Example-2: We consider 1, we target the second one subject to detect single case Outlier. Where v_{ii} is the $i-th$ diagonal element $\hat{V} = \widehat{W}^{1/2}X(X^T\widehat{W}X)^{-1}X^T\widehat{W}^{1/2}$.

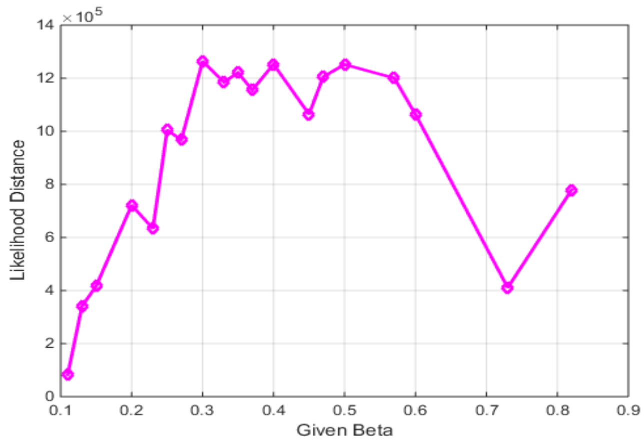

Figure-2: Plot for the Table-1 second individual for single case outlier under the model (15).

Example-3: We consider one another example to detect multiples Outliers cases,

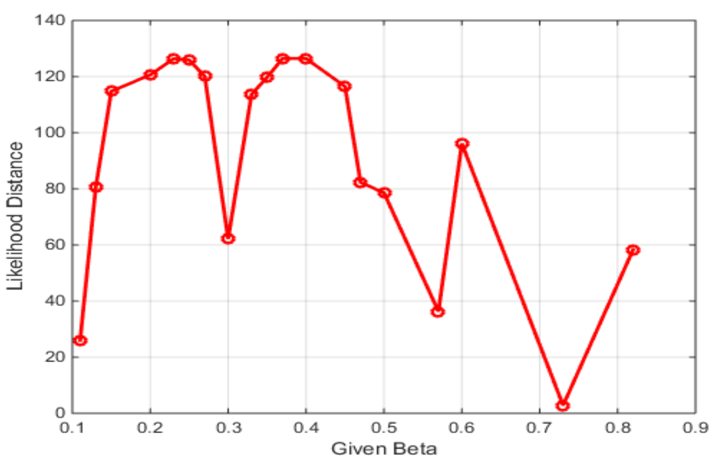

Figure-3: Plot for the Table-1 multiple outliers detection under the model (17).

Conclusion

While a few or small groups of observations are unique in few ways from most of the facts, the model fitting system may be significantly affected because all observations are forced into the same regression. Observations that make estimates deviate from that shape need to be recognized and omitted from the model fitting in influential cases. Its miles widely known that each observation does not play identical role in figuring out numerous consequences from a regression evaluation. As an example, the individual regression line is perhaps determined by some observations, whilst maximum of the information is mainly left out. Such perceptions that unimaginably impact the outcomes of the evaluation are known as powerful discourse. For regression evaluation, detection of outliers could be a very crucial step. The quantity of writing and research on effect evaluation for nonlinear relapse models isn't as monstrous as inside the straight case. In this paper, we have presented a Gauss-Newton method for parameter estimation and as properly we worked on rebut version of likelihood distance in single and multiple cases to detect outlier's data points for binomial response data.

References

1. Vonesh E. (1992). Nonlinear models for the analysis of longitudinal data. *Statistics in medicine*, 11(11), 1929-1954.
2. McKenzie E., Gettinby G., Lloyd J. and Caddy B. (1995). A statistical model for the response patterns to chemical tests for the absence or presence of trace materials. *Science & Justice*, 35(1), 31-36.
3. Pregibon D. (1981). Logistic regression diagnostics. *The Annals of Statistics*, 9(4), 705-724.
4. Bedrick E.J. and Tsai C.L. (1993). Diagnostics for binomial response models using power divergence statistics. *Computational statistics & data analysis*, 15(4), 381-392.
5. Altman N. and Krzywinski M. (2016). Erratum: Analyzing outliers: Influential or nuisance (Nature Methods). *Nature Methods*, 13(6), 281-282.
6. Gökpınar E. (2018). Standardized Log-Likelihood Ratio Test for the Equality of Inverse Gaussian Scale Parameters. *Iranian Journal of Science and Technology, Transactions A: Science*, 2(10), 1-7.
7. Vanegas L.H., Rondón L.M. and Cysneiros F.J.A. (2012). Diagnostic procedures in Birnbaum–Saunders nonlinear regression models. *Computational Statistics & Data Analysis*, 56(6), 1662-1680.
8. Cook R.D. (1979). Influential observations in linear regression. *Journal of the American Statistical Association*, 74(365), 169-174.
9. Ellenberg J.H. (1976). Testing for a single outlier from a general linear regression. *Biometrics*, 32(3), 637-645.
10. Kalivas J.H. (1999). Cyclic subspace regression with analysis of the hat matrix. *Chemometrics and intelligent laboratory systems*, 45(1-2), 215-224.
11. Chambers R.L. (1986). Outlier robust finite population estimation. *Journal of the American Statistical Association*, 81(396), 1063-1069.
12. Presscot C.R.D. (1981). Approximation significance levels for detecting outlier in linear regression. *Technimetrics*, 23, 59-64.
13. Noy D. and Menezes R. (2018). Parameter estimation of the Linear Phase Correction model by hierarchical linear models. *Journal of Mathematical Psychology*, 84, 1-12.

14. Ali M., Ali Z. and Choo A. (2018). Diagnostics of single and multiple outliers on likelihood distance. *AJER,* 07(7), 352-357.

15. Law M. and Jackson D. (2017). Residual plots for linear regression models with censored outcome data: A refined method for visualizing residual uncertainty. *Communications in Statistics-Simulation and Computation*, 46(4), 3159-3171.

16. Ong V.M.H., Nott D.J., Tran M.N., Sisson S.A. and Drovandi C.C. (2018). Likelihood-free inference in high dimensions with synthetic likelihood. *Computational Statistics & Data Analysis*, 128, 271-291.

17. Deunis Cook R. (1947). Sanford Veisberg "Residuals and influence in regression. ISBN 0-412-24280-X.

18. Schader M. and Schmid F. (1985). Computation of ML estimates for the parameters of a negative binomial distribution from grouped data. A comparison of the scoring, Newton—Raphson and E-M algorithms. *Applied Stochastic Models and Data Analysis*, 1(1), 11-23.

19. Inahama Y. (2010). A stochastic Taylor-like expansion in the rough path theory. *Journal of Theoretical Probability*, 23(3), 671-714.

20. Landwehr J.M., Pregibon D. and Shoemaker A.C. (1984). Graphical methods for assessing logistic regression models. *Journal of the American Statistical Association*, 79(385), 61-71.