Journal of the Korean
Statistical Society
Vol. 22, No. 1, 1993

On the Logistic Regression Diagnostics†

Choongrak Kim[1] and Kwang-Mo Jeong[2]

ABSTRACT

Since the analytic expression for a diagnostic in the logistic regression model is not available, one-step estimation is often used by a case-deletion point of view. In this paper, infinitesimal perturbation approach is used, and it is shown that the scale transformation of infinitesimal perturbation approach is eventually equal to the weighted perturbation of local influence approach and the replacement measure. Also, multiple cases deletion for the masking effect is considered.

KEYWORDS: Infinitesimal perturbation, Local influence, Logistic regression, Replacement measure.

1. INTRODUCTION

Consider a sample $\mathbf{y}^{\mathrm{T}} = (y_1, y_2, \ldots, y_n)$ of independent random variables such that y_j is binomially distributed $B(n_j, p_j)$ with n_j known and p_j unknown. The logistic regression model specifies the relationship

$$\eta_j = \mathrm{logit}(p_j) = \log(p_j/(1-p_j)) = \mathbf{x}_j^{\mathrm{T}}\boldsymbol{\beta}, \quad j = 1, 2, \ldots, n, \qquad (1.1)$$

[1] Department of Statistics, Pusan National University, Pusan, 609-735, Korea.
[2] Department of Statistics, Pusan National University, Pusan, 609-735, Korea.
† This research was supported by the Ministry of Education (BSRI-92-116)

where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are p-vectors of explanatory variables and $\boldsymbol{\beta}$ is an unknown parameter vector. The maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is often found using Newton's method since the likelihood equations are nonlinear in $\boldsymbol{\beta}$.

To see the influence of the ith observation on $\hat{\boldsymbol{\beta}}$, it is useful to evaluate $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}$, where $\hat{\boldsymbol{\beta}}_{(i)}$ is the estimate of $\boldsymbol{\beta}$ based on $n-1$ points without the ith observation. However, the analytic form for $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}$ is not available since iterative methods are required to obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{(i)}$. To overcome this difficulty, Pregibon (1981) suggested a one-step estimator $\hat{\boldsymbol{\beta}}^1_{(i)}$ of $\hat{\boldsymbol{\beta}}_{(i)}$ for the influence measure of Cook's distance (Cook, 1977) type. To be more specific, let $\hat{p}_j = \exp(\mathbf{x}_j^T\hat{\boldsymbol{\beta}})/\{1+\exp(\mathbf{x}_j^T\hat{\boldsymbol{\beta}})\}$ and let $\mathbf{V}$ be an $n \times n$ diagonal element with the jth diagonal $v_j = n_j\hat{p}_j(1-\hat{p}_j)$. Also, let $\mathbf{r}$ be an n-vector with the jth element $r_j = e_j/\sqrt{v_j}$, where $e_j = y_j - \hat{y}_j$. Pregibon (1981) showed that

$$\hat{\boldsymbol{\beta}}^1_{(i)} = \hat{\boldsymbol{\beta}} - \frac{(\mathbf{X}^T\mathbf{V}\mathbf{X})^{-1}\mathbf{x}_i e_i}{1-h^*_{ii}} \tag{1.2}$$

where h^*_{ii} is the ith diagonal element of $\mathbf{H}^* = \mathbf{V}^{1/2}\mathbf{X}(\mathbf{X}^T\mathbf{V}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{1/2}$. Pregibon (1981) discusses the accuracy of this one-step approximation and concludes that componentwise the approximation tends to underestimate the fully iterated value, but that this may be unimportant for identifying influential cases. Using (1.2), it can be easily shown that the one-step version of Cook's distance is

$$D_i = \frac{r_i^2}{p}\frac{h^*_{ii}}{(1-h^*_{ii})^2}. \tag{1.3}$$

Because of the masking effect, multiple cases deletion is required, and we can derive D_I where $\mathrm{I} = \{i_1, \ldots, i_m\}$ of a set of size m for the influential sets. Influential set version of (1.3) is

$$D_\mathrm{I} = \mathbf{r}_\mathrm{I}^T(\mathbf{I} - \mathbf{H}^*{}_\mathrm{I})^{-1}\mathbf{H}^*{}_\mathrm{I}(\mathbf{I} - \mathbf{H}^*{}_\mathrm{I})^{-1}\mathbf{r}_\mathrm{I}/p \tag{1.4}$$

where $\mathbf{r}_\mathrm{I}$ is m-subvector of $\mathbf{r}$ corresponding to cases in $\mathbf{I}$ and $\mathbf{H}^*{}_\mathrm{I}$ is $m \times m$ submatrix of $\mathbf{H}^*$. However, the computation of D_I is quite expensive as m increases.

In this paper, we consider an influence measure by the method of infinitesimal perturbation, while the Cook's distance is an influence measure by the deletion point of view. Also, we show that the infinitesimal perturbation scheme gives the same influence measure by the local influence (Cook, 1986) and the replacement measure (Kim; 1989,1992).

2. INFINITESSIMAL PERTURBATIONS, REPLACEMENT MEASURE, AND LOCAL INFLUENCE

2.1 Case Deletion

The infinitesimal perturbation approach in the multiple linear regression $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ is obtained by specifying $\epsilon_j \sim N(0, \sigma^2/w_j)$ where

$$w_j = \begin{cases} w & \text{if } j = i \\ 1 & \text{otherwise} \end{cases} \tag{2.1}$$

and $0 \leq w \leq 1$. Under this specification, the normal equations are $\mathbf{X}^{\mathrm{T}}\mathbf{W}(\mathbf{y} - \hat{\mathbf{y}}) = 0$ with $\mathbf{W} = diag(w_i)$, and the weighted least square estimator is $\hat{\boldsymbol{\beta}}_w = (\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{y}$. Pregibon (1981) showed that

$$\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_w = \frac{(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{x}_i(1-w)e_i}{\{1-(1-w)h_{ii}\}}. \tag{2.2}$$

The effect of infinitesimal perturbations of the variance of the ith data point is easily obtained by differentiation of $\hat{\boldsymbol{\beta}}_w$ leading to

$$\Delta\hat{\boldsymbol{\beta}}_w = \frac{\partial}{\partial w}\hat{\boldsymbol{\beta}}_w = \frac{(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{x}_i e_i}{\{1-(1-w)h_{ii}\}^2} \tag{2.3}$$

where $e_i = y_i - \hat{y}_i$ and h_{ii} is the ith diagonal element of $\mathbf{H} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}$. Evaluation at $w = 1$ describes local changes in $\hat{\boldsymbol{\beta}}_w$ at the usual least squares solution. In the literature of robust and resistant estimation, this function is termed the influence curve of the estimate $\hat{\boldsymbol{\beta}}$.

Kim (1989) proposed an influence measure called replacement measure defined as

$$R_i = (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{r(i)})^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{r(i)})/ps^2 \tag{2.4}$$

where $\hat{\boldsymbol{\beta}}_{r(i)} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}_r$, $\mathbf{y}_r = (y_1, \ldots, y_{i-1}, \hat{y}_i, y_{i+1}, \ldots, y_n)$, i.e., $\mathbf{y}_r$ is $\mathbf{y}$ with y_i replaced by $\hat{y}_i$, and $s^2 = \sum e_i^2/(n-p)$ is unbiased estimator of σ^2. To express R_i as a function of basic building blocks, let $\mathbf{N}_i$ be an $n \times n$ matrix with 1 in the ith diagonal element and zeros elsewhere, i.e., $\mathbf{N}_i = diag(0, \ldots, 0, 1, 0, \ldots, 0)$, then

$$\mathbf{y}_r = (\mathbf{I} - \mathbf{N}_i)\mathbf{y} + \mathbf{N}_i\hat{\mathbf{y}}$$

which gives

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{r(i)} &= (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\left[(\mathbf{I}-\mathbf{N}_i)\mathbf{y}+\mathbf{N}_i\hat{\mathbf{y}}\right] \\ &= \hat{\boldsymbol{\beta}} - (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{N}_i(\mathbf{I}-\mathbf{H})\mathbf{y} \\ &= \hat{\boldsymbol{\beta}} - (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{N}_i\mathbf{e}.\end{aligned}$$

Therefore,

$$\begin{aligned}R_i &= (\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{r(i)})^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{r(i)})/ps^2 \\ &= e_i^2 h_{ii}/ps^2.\end{aligned} \qquad (2.5)$$

We see that there is an interesting connection between R_i and $\Delta\hat{\boldsymbol{\beta}}_w$ in (2.3). Let $\Delta\hat{\boldsymbol{\beta}}_1$ be the value of $\Delta\hat{\boldsymbol{\beta}}_w$ evaluated at $w=1$, then it is easy to show that

$$R_i = \Delta\hat{\boldsymbol{\beta}}_1^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\Delta\hat{\boldsymbol{\beta}}_1/ps^2 \qquad (2.6)$$

i.e., R_i is a scalar version of influence curve evaluated at $w=1$.

Finally, consider the local influence proposed by Cook (1986). He suggested a general method for assessing the local influence of minor perturbations of a statistical model. More specifically, let $\boldsymbol{l}_{max}$ be a unit normed n-vector such that

$$\boldsymbol{l}^{\mathrm{T}}\mathbf{D_e}\mathbf{H}\mathbf{D_e}\boldsymbol{l}/ps^2$$

becomes maximum when $\boldsymbol{l} = \boldsymbol{l}_{max}$, where $\mathbf{D_e} = diag(e_i)$. Cook (1986) suggested to use $\boldsymbol{l}_{max}$ as diagnostic for the local influence of observations since $\boldsymbol{l}_{max}$ indicates how to perturbe the postulated model to obtain the greatest local change in the likelihood displacement. If $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ are eigenvalues of $\mathbf{D_e}\mathbf{H}\mathbf{D_e}$, then $\boldsymbol{l}_{max}$ is just the eigenvector corresponding to λ_1. Therefore, local influence is one way of perturbing $\mathbf{D_e}\mathbf{H}\mathbf{D_e}$ as $\boldsymbol{l}_{max}^{\mathrm{T}}\mathbf{D_e}\mathbf{H}\mathbf{D_e}\boldsymbol{l}_{max}$, and the amount of variation explained by $\boldsymbol{l}_{max}$ is just $\lambda_1/\sum_{i=1}^{n}\lambda_i$. By these respects, we consider a weighted perturbation of $\mathbf{D_e}\mathbf{H}\mathbf{D_e}$. To be more specific, we have by the eigenvalue decomposition theorem,

$$\mathbf{D_e}\mathbf{H}\mathbf{D_e} = \mathbf{P}\mathbf{D}_\lambda\mathbf{P}^{\mathrm{T}}$$

where $\mathbf{P}$ is an orthogonal matrix and $\mathbf{D}_\lambda = diag(\lambda_i)$. If we take as columns of $\mathbf{P}$ the eigenvectors of $\mathbf{D_e H D_e}$, it follows that

$$\mathbf{D_e H D_e} = \mathbf{P D_{\sqrt{\lambda}} D_{\sqrt{\lambda}} P^T}.$$

Let $\mathbf{L} = \mathbf{P D_{\sqrt{\lambda}}}$, then the columns of $\mathbf{L}$ reproduces $\mathbf{D_e H D_e}$ by the relation

$$\mathbf{D_e H D_e} = \mathbf{L L^T}$$

which implies that the ith diagonal element of $\mathbf{D_e H D_e}$ is equivalent to the i th diagonal element of $\lambda_1 \boldsymbol{l}_1 \boldsymbol{l}_1^T + \cdots + \lambda_n \boldsymbol{l}_n \boldsymbol{l}_n^T$, where $\boldsymbol{l}_1, \ldots, \boldsymbol{l}_n$ are eigenvectors corresponding to $\lambda_1, \ldots, \lambda_n$. Let

$$\mathbf{K} = \mathbf{D_e H D_e}/ps^2 \tag{2.7}$$

and define the influence of the ith observation is the ith diagonal element i.e., $\mathrm{K}(i) = \mathbf{K}_{ii} = e_i^2 h_{ii}/ps^2$. Conclusively, all the three measures mentioned above are equal to each other in the classical linear model.

We can extend the above argument to the logistic regression model. (Of course, we can extend to other generalized linear models like the log-linear model). In this case $\mathbf{K}$ in (2.7) is given by

$$\mathbf{K} = \mathbf{D_r H^* D_r}/p$$

where $\mathbf{D_r} = diag(r_i)$, and the case influence of the ith observation is

$$\mathrm{K}(i) = \mathbf{K}_{ii} = r_i^2 h_{ii}^*/p \tag{2.8}$$

which is also equivalent to R_i, the infinitesimal perturbation approach in Pregibon (1981). Analytic expression for the replacement measure is not available because $\hat{\boldsymbol{\beta}}_{r(i)}$ is not available , however, we verified that it is equal to (2.8) numerically.

2.2 Subset Deletion

We can extend three influence measures to the influence of a set of observations in $\mathrm{I} = \{i_1, \ldots, i_m\}$. First, consider the infinitesimal perturbation approach. For $m = 2$ with $i_1 = i, i_2 = j$, consider the multiple linear regression $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $\epsilon_k \sim N(0, \sigma^2/w_k)$ where

$$w_k = \begin{cases} w & \text{if } k = i \text{ or } j \\ 1 & \text{otherwise} \end{cases}$$

and $0 \leq w \leq 1$. Then, the normal equations are $\mathbf{X}^T\mathbf{W}(\mathbf{y}-\hat{\mathbf{y}}) = 0$ with $\mathbf{W} = diag(w_k)$, and $\hat{\boldsymbol{\beta}}_w = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y}$. Now,

$$\mathbf{X}^T\mathbf{W}\mathbf{X} = \mathbf{X}^T\mathbf{X} + (w-1)(\mathbf{x}_i\ \mathbf{x}_j)\begin{pmatrix} \mathbf{x}_i^T \\ \mathbf{x}_j^T \end{pmatrix}$$

and

$$\mathbf{X}^T\mathbf{W}\mathbf{y} = \mathbf{X}^T\mathbf{y} + (w-1)(\mathbf{x}_i\ \mathbf{x}_j)\begin{pmatrix} y_i \\ y_j \end{pmatrix}.$$

By using the updating formula in the Appendix A.2 in Cook and Weisberg (1982), we have

$$(\mathbf{X}^T\mathbf{X} + (w-1)(\mathbf{x}_i\ \mathbf{x}_j)\begin{pmatrix} \mathbf{x}_i^T \\ \mathbf{x}_j^T \end{pmatrix})^{-1}$$

$$= [(\mathbf{X}^T\mathbf{X})^{-1} + (1-w)(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{x}_i\ \mathbf{x}_j)[\mathbf{I}-(1-w)\mathbf{H}_\mathrm{I}]^{-1}\begin{pmatrix} \mathbf{x}_i^T \\ \mathbf{x}_j^T \end{pmatrix}(\mathbf{X}^T\mathbf{X})^{-1}]$$

where $\mathbf{H}_\mathrm{I} = \mathbf{X}_\mathrm{I}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}_\mathrm{I}^T$ with $\mathrm{I} = \{i,j\}$. Therefore,

$$\begin{aligned} \hat{\boldsymbol{\beta}}_w &= \hat{\boldsymbol{\beta}} + (\mathbf{X}^T\mathbf{X})^{-1}(w-1)(\mathbf{x}_i\ \mathbf{x}_j)\begin{pmatrix} y_i \\ y_j \end{pmatrix} + (1-w)(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{x}_i\ \mathbf{x}_j) \\ &\quad \cdot[\mathbf{I}-(1-w)\mathbf{H}_\mathrm{I}]^{-1}\begin{pmatrix} \mathbf{x}_i^T \\ \mathbf{x}_j^T \end{pmatrix}(\mathbf{X}^T\mathbf{X})^{-1}[\mathbf{X}^T\mathbf{y} - (1-w)(\mathbf{x}_i\ \mathbf{x}_j)\begin{pmatrix} y_i \\ y_j \end{pmatrix}] \end{aligned}$$

and

$$\begin{aligned} \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_w &= (1-w)(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{x}_i\ \mathbf{x}_j)\begin{pmatrix} y_i \\ y_j \end{pmatrix} - (1-w)(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{x}_i\ \mathbf{x}_j) \\ &\quad \cdot[\mathbf{I}-(1-w)\mathbf{H}_\mathrm{I}]^{-1}[\begin{pmatrix} \hat{y}_i \\ \hat{y}_j \end{pmatrix} - (1-w)\mathbf{H}_\mathrm{I}\begin{pmatrix} y_i \\ y_j \end{pmatrix}] \\ &= (1-w)(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{x}_i\ \mathbf{x}_j)[\mathbf{I}-(1-w)\mathbf{H}_\mathrm{I}]^{-1}\begin{pmatrix} e_i \\ e_j \end{pmatrix} \end{aligned}$$

which gives

$$\Delta\hat{\boldsymbol{\beta}}_w = \tfrac{\partial}{\partial w}\hat{\boldsymbol{\beta}}_w = -(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}(\mathbf{x}_i\ \ \mathbf{x}_j)[\mathbf{I}-(1-w)\mathbf{H}_{\mathrm{I}}]^{-1}\begin{pmatrix} e_i \\ e_j \end{pmatrix}$$

$$+(1-w)(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}(\mathbf{x}_i\ \ \mathbf{x}_j)\tfrac{\partial}{\partial w}[\mathbf{I}-(1-w)\mathbf{H}_{\mathrm{I}}]^{-1}\begin{pmatrix} e_i \\ e_j \end{pmatrix}$$

and the evaluation of $\Delta\hat{\boldsymbol{\beta}}_w$ at $w=1$ gives

$$\Delta\hat{\boldsymbol{\beta}}_1 = -(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}(\mathbf{x}_i\ \ \mathbf{x}_j)\begin{pmatrix} e_i \\ e_j \end{pmatrix}.$$

Therefore,

$$\begin{aligned} R_{ij} &= \Delta\hat{\boldsymbol{\beta}}_1^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\Delta\hat{\boldsymbol{\beta}}_1/ps^2 \\ &= (e_i^2h_{ii}+2e_ie_jh_{ij}+e_j^2h_{jj})/ps^2. \end{aligned} \tag{2.9}$$

Of course, we can extend (2.9) to the general case $\mathrm{I}=\{i_1,\ldots,i_m\}$. In fact,

$$R_{\mathrm{I}} = \{\sum_{i\in\mathrm{I}} e_i^2h_{ii} + 2\sum_{i\neq j;(i,j)\in\mathrm{I}} e_ie_jh_{ij}\}/ps^2. \tag{2.10}$$

Second, consider the replacement measure for $m=2$, let $\mathbf{N}_{ij}$ be a diagonal matrix with 1's in the ith and jth diagonal component and zeros elsewhere. Since $\hat{\boldsymbol{\beta}}_{r(i,j)} = \hat{\boldsymbol{\beta}} - (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{N}_{ij}\mathbf{e}$ we have

$$\begin{aligned} R_{ij} &= (\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{r(i,j)})^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{r(i,j)})/ps^2 \\ &= \mathbf{e}^{\mathrm{T}}\mathbf{N}_{ij}\mathbf{H}\mathbf{N}_{ij}\mathbf{e}/ps^2 \\ &= (e_i^2h_{ii}+2e_ie_jh_{ij}+e_j^2h_{jj})/ps^2 \end{aligned}$$

which is equivalent to (2.9). Also, it is quite clear for a set $\mathrm{I}=\{i_1,\ldots,i_m\}$ and we can get (2.10). Note that $R_{\mathrm{I}} \le \left(\sum_{i\in\mathrm{I}}\sqrt{R_i}\right)^2$ since $h_{ij}^2 \le h_{ii}h_{jj}$. Hence, an influential set should contain influential observations as its subset, i.e., the replacement measure is free of masking. Also, R_{I} is easier to compute than D_{I} in (1.5).

Finally, we consider, a weighted perturbation of $\mathbf{K} = \mathbf{D_e H D_e}/ps^2$ in (2.7). For a set I, we define

$$\mathrm{K(I)} = \sum_{i \in \mathrm{I}} \mathbf{K}_{ii} + 2 \sum_{i \neq j;(i,j) \in \mathrm{I}} \mathbf{K}_{ij}$$

where $\mathbf{K}_{ij}$ is the ijth element of $\mathbf{K}$. Therefore, three measures are equivalent to each other in the subset deletion case, too.

In the logistic regression model, a version of subset deletion is exactly the same as the argument given above, i.e.,

$$R_\mathrm{I} = \mathrm{K(I)} = \{\sum_{i \in \mathrm{I}} r_i^2 h_{ii}^* + 2 \sum_{i \neq j;(i,j) \in \mathrm{I}} r_i r_j h_{ij}^*\}/p.$$

3. EXAMPLE

As an illustrative example, we use Finney's (1947) data. The data, listed in Table 1, were obtained in a carefully controlled study of the effect of the rate and volume of air inspired on a transient vasoconstriction in the skin of the digits. Pregibon (1981) and Cook (1986) applied the logistic model

$$\mathrm{logit}(p) = \beta_0 + \beta_1 \log(\mathrm{RATE}) + \beta_2 \log(\mathrm{VOLUME})$$

and obtained one-step estimator by case-deletion point of view. We compute D_i and R_i for each observation. As is clear from Table 1, cases 4 and 18 are very influential. In addition, we compute and list four largest D_I and R_I for $m = 1,2,3$ in Table 2. When $m = 2$, a set (4,18) has large values of D_I and R_I, and other sets do not deserve special attention. When $m = 3$, no set draws special attention, and sets with large values of D_I and R_I are due to swamping by a subset (4,18). We can conclude that a set (4,18) is jointly influential.

Table 1. Finney's data with case influence based on the distance D_i and the replacement measure R_i

i	y	x_1	x_2	D_i	R_i
1	1	3.70	0.825	0.00183	0.00150
2	1	3.50	1.090	0.00028	0.00026
3	1	1.25	2.500	0.00197	0.00174
4	1	0.75	1.500	0.42894	0.35774
5	1	0.80	3.200	0.01379	0.01078
6	1	0.70	3.500	0.02620	0.01882
7	0	0.60	0.750	0.00000	0.00000
8	0	1.10	1.700	0.02176	0.01939
9	0	0.90	0.750	0.00011	0.00010
10	0	0.90	0.450	0.00000	0.00000
11	0	0.80	0.570	0.00000	0.00000
12	0	0.55	2.750	0.01753	0.01272
13	0	0.60	3.000	0.04655	0.03263
14	1	1.40	2.330	0.00134	0.00120
15	1	0.75	3.750	0.01122	0.00842
16	1	2.30	1.640	0.00036	0.00033
17	1	3.20	1.600	0.00003	0.00003
18	1	0.85	1.415	0.32804	0.26844
19	0	1.70	1.060	0.06679	0.05038
20	1	1.80	1.800	0.00113	0.00101
21	0	0.40	2.000	0.00016	0.00014
22	0	0.95	1.360	0.00737	0.00595
23	0	1.35	1.350	0.03118	0.02661
24	0	1.50	1.360	0.05197	0.04478
25	1	1.60	1.780	0.00247	0.00219
26	0	0.60	1.500	0.00052	0.00046
27	1	1.80	1.500	0.00336	0.00293
28	0	0.95	1.900	0.01992	0.01743
29	1	1.90	0.950	0.06536	0.04522
30	0	1.60	0.400	0.00018	0.00017
31	1	2.70	0.750	0.05539	0.03149
32	0	2.35	0.030	0.00000	0.00000
33	0	1.10	1.830	0.02750	0.02476
34	1	1.10	2.200	0.00673	0.00594
35	1	1.20	2.000	0.00601	0.00537
36	1	0.80	3.330	0.01173	0.00913
37	0	0.95	1.900	0.01992	0.01743
38	0	0.75	1.900	0.00979	0.00793

Table 2. Four largest sets based on D_{I} and R_{I} for $m = 1, 2, 3$ in the Finney's data

m	set	D_{I}	set	R_{I}
1	4	.429	4	.358
	18	.328	18	.268
	19	.067	19	.050
	29	.065	29	.045
2	4,18	1.856	4,18	1.243
	4,29	.650	4.29	.481
	18,29	.579	18,29	.403
	4,31	.493	4,31	.393
3	4,18,29	2.408	4,18,29	1.455
	4,18,22	2.084	4,18,31	1.301
	4,18,26	2.043	4,18,39	1.292
	4,18,38	2.000	4,18,32	1.243

4. CONCLUDING REMARKS AND FUTURE RESEARCH

In the generalized linear models including logistic regression model, analytic expression for $\hat{\boldsymbol{\beta}}$ is not available. Therefore, one-step estimator is often used. However, they underestimate the full iterated value. In this paper we suggest a replacement measure which has interesting connections with the influence curve and local influence. This measure has much simpler form than Cook's distance and is computationally easier.

For the future research areas, we raise two issues. First, some cutoff values for influential observations are needed. Those values are functions of the number of observations n, the number of covariates p, and the number of cases m. It can be done by the simulation, however, it requires a lot of computation and a careful choice of design matrix. Second, geometric justification for the equivalence of three measures will be very useful because those measures are originated from the different point of view.

REFERENCE

(1) Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, 19, 15-18.

(2) Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society*, Ser B, 48, 133-169.

(3) Finney, D. J. (1947). The estimation from individual records of the relationship between dose and quantal response. *Biometrika*, 34, 320-334.

(4) Kim, C. (1989). A study on influential sets in regression. Unpublished Ph. D. Thesis, Univ. of Wisconsin-Madison.

(5) Kim, C. (1992). Local influence and replacement measure (manuscript).

(6) Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics*, bf 9, 705-724.