The Korean Communications
in Statistics Vol. 9, No. 2, 2002
pp. 521-531

A Study on Detection of Influential Observations on A Subset of Regression Parameters in Multiple Regression[1]

Sung Hyun Park[2] and Jin Ho Oh[3]

Abstract

Various diagnostic techniques for identifying influential observations are mostly based on the deletion of a single observation. While such techniques can satisfactorily identify influential observations in many cases, they will not always be successful because of some mask effect. It is necessary, therefore, to develop techniques that examine the potentially influential effects of a subset of observations. The partial regression plots can be used to examine an influential observation for a single parameter in multiple linear regression. However, it is often desirable to detect influential observations for a subset of regression parameters when interest centers on a selected subset of independent variables. Thus, we propose a diagnostic measure which deals with detecting influential observations on a subset of regression parameters. In this paper, we propose a measure *M*, which can be effectively used for the detection of influential observations on a subset of regression parameters in multiple linear regression. An illustrated example is given to show how we can use the new measure *M* to identify influential observations on a subset of regression parameters.

Keywords : Influential observation, Mask effect, Partial regression plots

1. Introduction

Assume the usual linear model $y = X\beta + \varepsilon$, where y is an $n\times 1$ vector of observed responses, X is an $n\times p$ design matrix of full rank, β is the $p\times 1$ vector of unknown regression parameters, and ε is the $n\times 1$ error vector where ε_i is

1) This research work of Sung Hyun Park was supported in part by KOSEF through Statistical Research Center for Complex Systems at Seoul National University, and the research work of Jin Ho Oh was partially supported by the Brain Korea 21 Project.

2) Professor, Department of Statistics, Seoul National University, 151-742, Korea.
E-mail:parksh@plaza.snu.ac.kr

3) Department of Statistics, Seoul National University, Seoul, 151-742, Korea.

identically and independently distributed with mean 0 and variance σ^2. Note that $\boldsymbol{\beta}$ is estimated by the least squares estimator $\boldsymbol{b}=(X'X)^{-1}X'\boldsymbol{y}$. Then the residual vector $\boldsymbol{e}$ becomes $\boldsymbol{e}=(I-X(X'X)^{-1}X')\boldsymbol{y}$.

First, we present several methods that deal with detecting influential observation(s) on a subset of regression parameters i.e partial regression plot, D_i'(proposed by Park & Wakimoto (1987)), $D_i(G)$(proposed by Cook & Weisberg (1980, 1982)) and M(the measure proposed in this paper). Second, we will illustrate an example to show how these measures perform in multiple regression. For some good references of regression diagnostics, see Belsley, Kuh and Welsch (1980), Cook(1977, 2000) and Kang and Park(1988).

2. Detection of Influential Observations

One outlier can mask the effect of another. Therefore, to develop techniques that examine the potentially influential effects of a subset of observations, a natural multiple-row generalization of $b_j-b_{j(i)}$ would be desirable, where $b_{j(i)}$ is the least squares estimate of β_j when the i^{th} observation is deleted. A generalization is to examine larger values of $|b_j-b_{j(D_m)}|/scale$ for j=1, ... , p and m=2, 3, 4 so on, where ***'scale'*** indicates an appropriate measure of standard error. Here D_m is a set(of size m) of indexes of the rows to be deleted and $b_{j(D_m)}$ is the least-squares estimate of β_j when the subset of observation is deleted.

If fitted values are of interest, then the appropriate measure becomes

$$\frac{|\boldsymbol{x}_k'[\boldsymbol{b}-\boldsymbol{b}_{(D_m)}]|}{scale} \tag{1}$$

for k=1, ... , n, when $\boldsymbol{x}_k$ is the k^{th} row of design matrix X. Although computational formulas exist for these quantities(Bingham 1977), the cost is great and we feel most of the benefits can be obtained more simply.

To avoid the consideration of p quantities in $|b_j-b_{j(D_m)}|/scale$ or n quantities in **(1)**, squared norms, such as

$$[\boldsymbol{b}-\boldsymbol{b}_{(D_m)}]'X'X[\boldsymbol{b}-\boldsymbol{b}_{(D_m)}] \tag{2}$$

can be considered as a summary measure. Since we are often most interested in changes in fit that occur for the data points retaining after deletion, **(2)** can be modified to

$$MDFFIT=[\boldsymbol{b}-\boldsymbol{b}_{(D_m)}]'X'_{(D_m)}X_{(D_m)}[\boldsymbol{b}-\boldsymbol{b}_{(D_m)}] \tag{3}$$

which was suggested by Bingham(1977). He also shows that (3) can be expressed as

$$MDFFIT = e'_{D_m} X_{D_m} [X'_{(D_m)} X_{(D_m)}]^{-1} X'_{D_m} e_{D_m} \tag{4}$$

where, e_{D_m} is the column vector of least-squares residuals for the rows whose indexes are contained in D_m.

The situation in which a subset of regression parameters is of special interest may occur in practice. For this case, the partial regression model can be redefined to be used for identifying an influential observation on a selected subset of regression parameters.

Suppose the partition

$X = (X_A : X_B) = (x_1, x_2, \cdots, x_{p-q} : x_{p-q+1}, \cdots, x_p)$, $\beta' = (\beta'_1, \beta'_2)$

where X_A is $n \times (p-q)$, X_B is $n \times q$, β_1 is $(p-q) \times 1$ and β_2 is $q \times 1$ matrices. Let b_1 and b_2 be the least squares estimates of β_1 and β_2 for the usual linear model $y = X\beta + \varepsilon$, Let us define as follows,

$e_{y|-[B]}$: the residual vector from the least squares regression of y on X_A where the regressors indexed $[B] = (p-q+1, \cdots, p)$ are omitted.

$x_{j \cdot [B]}$: the residual vector from the least squares regression of x_j on X_A where x_j is a column vector of X_B.

$X_{.[B]}$: the residual matrix where its columns are in the order of $x_{j \cdot [B]}$, $j = p-q+1, \cdots, p$ and then $X_{.[B]} = [I - X_A(X'_A X_A)^{-1} X'_A] X_B$.

Consider the multiple regression model between $e_{y|-[B]}$ and $X_{.[B]}$,

$$e_{y|-[B]} = X_{.[B]} \beta^*_2 + \varepsilon^* \tag{5}$$

This is a multiple regression model between y and x_{p-q+1} , ⋯ , x_p which are both adjusted for x_1 , ⋯ , x_{p-q} by regression relationship. Park and Wakimoto(1987) showed that the least squares estimate of β^*_2 is b_2, and the resulting residual vector is in fact equal to e. The rationale for (5) is that the partial regression model with respect to the last q components of β can be considered as an adjusted model for the effect of the first p-q regressors.

Consider the partial regression model (5). If we let the i^{th} element of $e_{y|-[B]}$ be $e_{y|-[B]i}$ and its predicted value be $\widehat{e_{y|-[B]i}}$. Then the proposed statistic in Park and Wakimoto is given by

$$D_i' = \frac{(r'_i)^2 h_i}{q(1-h_i)} \tag{6}$$

where $r'_i = \dfrac{(e_{y|-[B]} - \widehat{e_{y|-[B]i}})}{s'\sqrt{1-h_i}}$

h_i : the i^{th} diagonal element of $X_{.[B]}(X'_{.[B]}X_{.[B]})^{-1}X'_{.[B]}$.

s' : the root of the residual mean squared error from the partial regression model.

Also note that since $s^2 = \dfrac{e'e}{(n-p)}$, $(s')^2 = \dfrac{e'e}{n-q}$, and $(s')^2 = s^2\dfrac{(n-p)}{n-q}$, s' is slightly smaller than s.

The computation of D_i' appears to be difficult. However, since $X_{.[B]} = [I - X_A(X'_A X_A)^{-1}X'_A]X_B$, we can obtain the following equality relationship $X(X'X)^{-1}X' = X_A(X'_AX_A)^{-1}X'_A + X_{.[B]}(X'_{.[B]}X_{.[B]})^{-1}X'_{.[B]}$, which enables us to have

$$v_i = u_i + h_i \tag{7}$$

which v_i is the ith diagonal element of $X(X'X)^{-1}X'$ and u_i is the ith diagonal element of $X_A(X'_AX_A)^{-1}X'_A$. The more generalized form of the equation (7) is given in Rao and Yanai(1979). Therefore, by applying this result, we can get the following statistics

$$D_i' = \frac{(r_i')^2(v_i - u_i)}{q(1-v_i+u_i)} \tag{8}$$

Cook and Weisberg(1980) proposed $D_i(G) = \dfrac{(r_i')^2(v_i-u_i)}{q(1-v_i)}$ to detect a simple influential observation on a selected subset of β. From the definitions of D_i' and $D_i(G)$, we can readily show that

$$D_i(G) = D_i'\left(\frac{1-v_i+u_i}{1-v_i}\right)^2\left(\frac{n-q}{n-p}\right) \tag{9}$$

3. Proposition of a measure

In this section, we want to propose a measure that deals with detecting influential observations on a subset of regression parameters. For this, we represent the formula (2) in terms of Sb which are an extension of the previous discussion to accommodate the situation in which q linearly independent combinations of the elements of β are of interest. For instance, if the last q component of β are of interest, then S becomes (O, I_q). where O is the matrix of all zero elements and I_q is the $q\times q$ identity matrix. A composite measure of influence of the m observations on the coefficients in Sb must be expressed as a scalar quantity that somehow standardizes $d_m = Sb - Sb_{(D_m)}$. A proper standardization is to produce a quadratic form in which d_m is standardized by inverse of the variance-covariance matrix. Note that $Var(Sb)$ is

$$S(X'X)^{-1}S'\sigma^2.$$

Thus, we propose M which deals with detecting influential observations on a subset of regression parameters as follows.

$$\begin{aligned} M &= [Sb - Sb_{(D_m)}]' \,(S(X'X)^{-1}S')^{-1}[Sb - Sb_{(D_m)}] \\ &= [b - b_{(D_m)}]' \,Q_1\, [b - b_{(D_m)}] \end{aligned} \tag{10}$$

where, $Q_1 = S'[S(X'X)^{-1}S']^{-1}S$.

The quantity in equation **(10)** becomes a standardized version of d_m. For helping to understand M, a convenient formula for $b - b_{(D_m)}$ can be derived from the result (Bingham 1977)

$$b - b_{(D_m)} = -(X'X)^{-1}X'_{D_m}[I - V_{D_m}]^{-1}r_{D_m} \tag{11}$$

Here, the subscript ' (D_m)' means "with the m cases indexed by m deleted" , while D_m without parentheses means "with only the cases indexed by m remaining". Here $V_{D_m} = X_{D_m}(X'_{D_m}X_{D_m})^{-1}X'_{D_m}$ and V_{D_m} can be decomposed into $V_{D_m} = P'\Lambda P$ where P is an orthogonal matrix and $\Lambda = diag(\lambda_1, \cdots, \lambda_q)$ with $0 \le \lambda_1 \le \cdots \le \lambda_q \le 1$ in which λ_i is an eigenvalue of V_{D_m}. Also, r_{D_m} is the vector of residuals for the cases indexed in D_m . For convenience, we suppress all indications that P and Λ depend on m.

The relation between the proposed statistic $\boldsymbol{M}$ and the equation (2) can be found as follows.

$$[b-b_{(D_m)}]'X'X[b-b_{(D_m)}]$$
$$=r'_{D_m}(I-V_{D_m})^{-1}X_{D_m}(X'X)^{-1}X'X(X'X)^{-1}X'_{D_m}(I-V_{D_m})^{-1}r_{D_m}$$
$$=r'_{D_m}(I-V_{D_m})^{-1}X_{D_m}(X'X)^{-1}X'_{D_m}(I-V_{D_m})^{-1}r_{D_m}$$

and,

$$\begin{aligned} M &= r'_{D_m}(I-V_{D_m})^{-1}X_{D_m}(X'X)^{-1}Q_1(X'X)^{-1}X'_{D_m}(I-V_{D_m})^{-1}r_{D_m} \\ &= (2)-r'_{D_m}(I-V_{D_m})^{-1}U_{D_m}(I-V_{D_m})^{-1}r_{D_m} \\ &= r'_{D_m}(I-V_{D_m})^{-1}(V_{D_m}-U_{D_m})(I-V_{D_m})^{-1}r_{D_m} \end{aligned} \qquad (12)$$

$$where, \begin{cases} (X'X)^{-1}Q_1(X'X)^{-1}=(X'X)^{-1}-\begin{pmatrix} (X'_{D_m}X_{D_m})^{-1} & 0 \\ 0 & 0 \end{pmatrix} \\ U_{D_m}=X_{D_m}(X'_{D_m}X_{D_m})^{-1}X'_{D_m} \end{cases}$$

Thus, when a single case is deleted(m=1) $M_{(m=1)}=r_i^2(v_i-u_i)/(1-v_i)^2$ from **(12)**. The influence of a single case on a selected subset of β may therefore be determined from the result of two separate regressions on the full data.

The relation among the D_i', $D_i(G)$ and $M_{(m=1)}$ can be also found as follows. From the definition of D_i', $D_i(G)$ and $M_{(m=1)}$, we can show the following relationship,

$$M_{(m=1)}=\frac{D_i(G)\times q}{(1-v_i)},\quad D_i(G)=D_i'\left(\frac{1-v_i+u_i}{1-v_i}\right)^2\frac{(n-q)}{(n-p)} \qquad (13)$$

$$\text{where,} \begin{cases} D_i' =\dfrac{(r_i')^2h_i}{q(1-h_i)},\quad D_i(G)=\dfrac{(r_i')^2(v_i-u_i)}{q(1-v_i)} \\ M_{(m=1)}=\dfrac{r_i^2(v_i-u_i)}{(1-v_i)^2},\quad \textit{a single case of } M \end{cases}$$

These results imply that $M_{(m=1)}$ is more influenced by q and $1-v_i$ as compared with $D_i(G)$, and $D_i(G)$ is more influenced by u_i as compared with D_i'. But $M_{(m=1)}$ is easier for computation than D_i' and $D_i(G)$. In this point of view, $M_{(m=1)}$ could be better than D_i' and $D_i(G)$ for practical use.

4. An Illustrated Example

Until now we presented detection of influential observations on a subset of regression parameters in multiple regression. In this section, we will illustrate an example to show how these statistics play roles in multiple regression. As a device of obtaining these statistics, we used the statistical package SAS 8.01.

4.1. Data and Analysis

The data set (Table 1) is given in Daniel and Wood (1971). y is change of rut depth in inches/million wheel passes, x_1 is viscosity of asphalt, x_2 is percentage of asphalt in surface course, x_3 is percentage of asphalt in base course, x_4 is indicator variable to separate two sets of runs, x_5 is percentage of fines in surface course, and x_6 is percentage of voids in surface course. It is suggested that the model is

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \varepsilon_i .$$

Table 1. Data set

number	x_1	x_2	x_3	x_4	x_5	x_6	y
1	2.8	4.68	4.87	-1	8.4	4.916	6.75
2	1.4	5.19	4.5	-1	6.5	4.563	13
3	1.4	4.82	4.73	-1	7.9	5.321	14.75
4	3.3	4.85	4.76	-1	8.3	4.865	12.6
5	1.7	4.86	4.95	-1	8.4	3.776	8.25
6	2.9	5.16	4.45	-1	7.4	4.397	10.67
7	3.7	4.82	5.05	-1	6.8	4.867	7.28
8	0.92	4.86	4.7	-1	8.6	4.828	12.67
9	0.68	4.78	4.84	-1	6.7	4.865	12.58
10	0.68	5.16	4.76	-1	7.7	4.034	20.6
11	6	4.57	4.82	-1	7.4	5.45	3.58
12	4.3	4.61	4.65	-1	6.7	4.853	7
13	0.6	5.07	5.1	-1	7.5	4.257	26.2
14	1.8	4.66	5.09	-1	8.2	5.144	11.67
15	6	5.42	4.41	-1	5.8	3.718	7.67
16	4.4	5.01	4.74	-1	7.1	4.715	12.25
17	88	4.97	4.66	1	6.5	4.625	0.76
18	62	5.01	4.72	1	8	4.977	1.35
19	50	4.96	4.9	1	6.8	4.322	1.44
20	58	5.2	4.7	1	8.2	5.087	1.6
21	90	4.8	4.6	1	6.6	5.971	1.1
22	66	4.98	4.69	1	6.4	4.647	0.85
23	140	5.35	4.76	1	7.3	5.115	1.2
24	240	5.04	4.8	1	7.8	5.939	0.56
25	420	4.8	4.8	1	7.4	5.916	0.72
26	500	4.83	4.6	1	6.7	5.471	0.47
27	180	4.66	4.72	1	7.2	4.602	0.33
28	270	4.67	4.5	1	6.3	5.043	0.26
29	170	4.72	4.7	1	6.8	5.075	0.76
30	98	5	5.07	1	7.2	4.334	0.8
31	35	4.7	4.8	1	7.7	5.705	2.00

Table 2 summarizes the results for these cases. In this table, the superscript "***" indicates the most influential observation, and "**" indicates the next most influential observation, and "*" indicates the third influential observation in the columns of Residual and DFFITS. The next potentially influential points that are so chosen are also given in Table 2.

In Table 3, the pair (10, 13) appears to be the most influential observations. As we can see from Table 2, the observation 10 is an outlier, since its residual is noticeably large. However, its h_{ii} and DFFITS are not large enough. Hence, as a single observation, it is not judged as influential. Thus we can observe that the observation 13 has substantially masked the impact of the observation 10.

Table 2. Influential observations

Obs	Residual	h_{ii}	DFFITS
1	-4.2751	0.1514	-0.5021
	⋮	⋮	⋮
5	-4.9006	0.3256	**-1.0842***
6	-0.9588	0.2197	-0.1434
7	-5.0809	0.2762	**-0.9647**
	⋮	⋮	⋮
10	**6.8048****	0.1491	0.8310
11	**-5.6528***	0.1877	-0.7929
12	-1.2432	0.2039	-0.1758
13	**11.1319*****	0.2610	**2.5823*****
14	-0.4065	0.1985	-0.0562
15	-4.9913	0.4173	**-1.4627****
	⋮	⋮	⋮
26	0.2584	**0.4881***	0.0877
	⋮	⋮	⋮
31	2.4863	0.2296	0.3887

* **the cutoff values** : $h_{ii}=0.45$, $|DFFITS|=0.95$

Table 3. *MDFFIT* of two observations

Joint observation	10,13 Obs.	9,13 Obs	11,13 Obs.	13,15 Obs	13,14 Obs
Value	81.712985	53.048504	49.546224	48.216902	47.025472

We want to detect the most influential observations for the following cases when two particular parameters from $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ and β_6 are of special interest. Case 1 is the case when of β_1 and one other parameter are of interest, and so on. Thus Case 1 ~ Case 6 cover all the cases involving any two parameters.

Case 1 : $(\beta_1, \beta_2), (\beta_1, \beta_3), (\beta_1, \beta_4), (\beta_1, \beta_5)$ and (β_1, β_6) are of interest.
Case 2 : $(\beta_2, \beta_3), (\beta_2, \beta_4), (\beta_2, \beta_5)$ and (β_2, β_6) are of interest.
Case 3 : $(\beta_3, \beta_4), (\beta_3, \beta_5)$ and (β_3, β_6) are of interest.

Case 4 : (β_4, β_5) and (β_4, β_6) are of interest.
Case 5 : (β_5, β_6) is of interest.
Table 4 summarizes the results for these cases.

Table 4. *M* values for each pair of observations for a subset of parameters

β_1, β_2 (*M value*)	β_1, β_3 (*M value*)	β_1, β_4 (*M value*)	β_1, β_5 (*M value*)	β_1, β_6 (*M value*)
①10, 13 (26.35)	①13, 15 (49.37)	① 7, 11 (18.21)	① 5, 21 (25.32)	① 5, 21 (44.73)
②15, 23 (22.13)	② 9, 13 (48.58)	②10, 13 (14.54)	② 7, 15 (23.31)	② 5, 30 (21.92)
③11, 13 (21.66)	③ 6, 13 (48.10)	③ 5, 21 (12.63)	③ 1, 5 (17.31)	③ 5, 31 (18.51)
④12, 13 (16.44)	④13, 14 (47.51)	④ 7, 15 (12.54)	④ 7, 11 (15.07)	④ 1, 5 (17.55)
⑤1, 13 (16.39)	⑤10, 13 (46.51)	⑤ 9, 13 (10.84)	⑤ 7, 10 (12.61)	⑤ 3, 5 (17.13)

β_2, β_3 (*M value*)	β_2, β_4 (*M value*)	β_2, β_5 (*M value*)	β_2, β_6 (*M value*)	β_3, β_4 (*M value*)
①10, 13 (50.20)	①10, 13 (28.67)	① 7, 15 (27.70)	①10, 13 (44.62)	① 9, 13 (53.78)
②13, 15 (50.16)	②15, 23 (22.18)	②15, 23 (25.56)	② 5, 21 (43.81)	②10, 13 (51.16)
③ 9, 13 (49.13)	③11, 13 (18.25)	③ 5, 13 (22.69)	③11, 13 (35.03)	③13, 14 (50.71)
④ 6, 13 (48.29)	④ 5, 13 (17.18)	④ 1, 5 (21.87)	④15, 23 (27.32)	④ 6, 13 (49.51)
⑤13, 14 (47.88)	⑤ 2, 13 (16.99)	⑤10, 13 (21.24)	⑤ 5, 15 (25.12)	⑤13, 15 (49.28)

β_3, β_5 (*M value*)	β_3, β_6 (*M value*)	β_4, β_5 (*M value*)	β_4, β_6 (*M value*)	β_5, β_6 (*M value*)
①13, 15 (59.44)	①10, 13 (54.93)	① 7, 15 (27.76)	① 5, 21 (48.51)	① 5, 21 (47.38)
② 9, 13 (48.26)	②13, 15 (50.12)	② 7, 11 (20.07)	②10, 13 (24.75)	② 1, 5 (26.39)
③13, 14 (48.26)	③ 6, 13 (50.01)	③ 5, 21 (19.76)	③ 5, 15 (23.19)	③ 7, 15 (23.89)
④ 6, 13 (48.19)	④ 9, 13 (49.93)	④ 1, 5 (17.86)	④ 5, 30 (21.92)	④ 5, 30 (22.49)
⑤10, 13 (48.18)	⑤13, 14 (49.14)	⑤ 9, 13 (16.29)	⑤ 5, 31 (21.22)	⑤ 7, 11 (18.65)

In Table 4, "①" indicates the most joint influential observations for each subset of regression parameters. And "②, ③, ④, ⑤" indicate the orderly influential joint observations for each subset of regression parameters.

First, the joint pair of observations (10, 13) is the most influential observations on (β_1, β_2), (β_2, β_3), (β_2, β_4), (β_2, β_6), and (β_3, β_6) of regression parameters. Second, the joint pair of observation (5, 21) is the most influential observations on (β_1, β_5), (β_1, β_6), (β_4, β_6), and (β_5, β_6). Third, the joint pair of observations (13, 15) is the most influential observations on (β_1, β_3) and (β_3, β_5). Fourth, the joint pair of observations (7, 15) is the most influential observations on (β_2, β_5) and (β_4, β_5). And, it seems that the observation 7 has substantially masked the impact of the observation 11 on (β_1, β_4).

In comparison with the results in Table 2 and Table 3 for all parameters, Table 4 for a subset of particular parameters gives us much more information about influential observations. Tables 4 provides different pairs of observations as the pairs of parameters of interest change.

5. Conclusion

In many situations, very often, we want to find out an influential subset of observations instead of only a single observation. Also it is of interest to find influential observations for some subsets of parameters. In these cases, M-statistic turns out to be a very useful measure, because it can give the best subset of observations for any combination of parameters.

Acknowledgements

The authors would like to express their thanks for the referees' comments that help to improve this paper very much. Their comments were very useful to make this paper clear and readable.

References

[1] Belsley, D.A., Kuh, E., and Welsch, R. E. (1980), Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, New York : Wiley.

[2] Bingham, Cristopher (1977), *"Some Identities Useful in the Analysis of Residual from Linear Regression,"* Technical Report 300, School of Statistics, University of Minnesota.

[3] Cook, R. D. (2000), " *Detection of influential observation in linear regression*" , *Technometrics,* 42(1), 65-68.

[4] Cook, R. D. and Weisberg, S. (1982), Residual and Influence in Regression, New York: Chapman & Hall.

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

[6] Cook, R. D. and Weisberg, S. (1980). "*Characterizations of an empirical influence function for detecting influential cases in regression.*" Technometrics 22, 495-508.

[7] Daniel, C. and Wood, F. S. (1971), "*Fitting Equations to Data*", Computer Analysis of Multifactor Data, New York: John Wiley & Sons.

[8] Kang, E. M. and Park, S. H. (1988), "*A study on detection of outliers and influential observations in linear models*" , Journal of the KSQC Vol. 16, No. 2, 18-33.

[9] Park, S. H. and Kazumasa, Wakimoto. (1987), "*Detection of an influential observation on a subset of regression parameters in multiple regression*" , JAPAN STATIST. SOC. Vol. 17, No. 1, 11-19.

[10] Rao, C. R. and Yanai, H. (1979). General definition and decomposition of projectors and some application to statistical problems. *Journal of Statistical Planning and Inference* **3**, 1-17.

[2002년 4월 접수, 2002년 8월 채택]