工業經營學會誌 第15卷 第25輯 1992年 5月 30日 103

A Comparision of Diagnostic Measures in Linear Regression
—회귀진단을 위한 새로운 척도의 제안 및 상호비교

최 성 운*

ABSTRACT

This paper is to study the various diagnostic measures for detecting outliers and influential cases in linear regression. In this paper we review the most common diagnostic measures and show the interrelationships the exist among them. Based on the PRESS(Predicted REsidual Sum of Squares) offered by Allen(1974) as a criterion for model selection, we propose three measures for detecting outliers and influential cases. Examples are given illustrating various diagnostic measures including proposed measures.

1. INTRODUCTION

Consider the standard linear regression model

$$Y = X\beta + \varepsilon \tag{1}$$

where Y is an n×1 vector of observable responses (values of response variables, dependent variables), X is an n×p' full rank matrix of known constants (values of explanatory variables, independent variables, predictors, regressors, carriers, factors) possibly including one constant explanatory variables, β is a p'×1 vector of unknown coefficients (parameters) to be estimated, and ε is an n×1 vector of unobservable errors which are normally and independently distributed with zero mean and unknown common variance σ^2.

Assuming the model (1) with rank(X)=P' and Cov(e)=σ^2 I, the least squares estimator of β using the full data $\hat{\beta} = (X^T X^{-1}) X^T Y$ and the full sample estimate of σ^2(residual mean square) is $\hat{\sigma}^2 = \frac{e^T e}{n-p'}$,

where e denote the vector of residuals, $e = Y - \hat{Y}$

We shall need additional notation. A subscript "(i)" added to a quantity means "with the ith case deleted". Thus, for example, $X_{(i)}$ is an (n−1)×p' matrix derived from X by deleting the ith row X_I^T, $\hat{\beta}_{(i)} = (X_{(i)}{}^T X_{(i)})^{-1} X_{(i)}{}^T Y_{(i)}$, and so on.

Also of importance is the hat(projection) matrix $H = (h_{ij}) = X(X^T X)^{-1} X^T$, an n×n rank p matrix that projects onto the column space of X. The diagonal entries h_{ii} are of special interest.

An influential case is one that, if removed, would substantially change certain important features of the regression analysis under consideration. For example, the deletion of a case may result in large changes in the estimated coefficients, the fitted values, or the estimated variances of parameters, and/or the goodness-of-fit statistics.

Regression diagnostics measure outlyingness (compatability of the dependent varible with the regression line fit to all cases), leverage (potential for drawing the regression line toward that case), and influence (effect that deletion of the case has on the fitted coefficients). Diagnostic measures for detecting outliers and influential cases are various. We review the inter-relationships that exist among them. Based on the PRESS as a criterion for model selection, we propose three measures for detecting outlirers and influential cases.

*Dept. of Industrial Engineering, Kyungwon University, Sungnam, Korea.

접수 : 1992. 4. 25.
확정 : 1992. 5. 2.

These measures can be classified as follows :

1. Diagnosic measures for *detecting outliers*
   - Outlyingness measures based on residuals,
2. Diagnostic measures for *detecting influential cases*
   - Influence measures based on the remoteness of points in the x-y space,
   - Influence measures based on the *sample influence curve*,
   - Influence measures based on the volume of confidence ellipsoids,
   - Influence measures based on the *likelihood distance*,
3. Diagnostic measures for detecting outliers and influential case,
   - Proposed measures based on the PRESS.

2. DIAGNOSTIC MEASURES FOR DETECTING OUTLIERS

2.1 Outlyingness Measures Based On Residuals

2.1.1 Ordinary Residuals

$$e_i = y_i - \hat{y}_i, \quad i=1, 2, \cdots, n \tag{2}$$

2.1.2 Internally Studentized Residuals

$$r_i = \frac{e_i}{\hat{\sigma}\sqrt{1-h_{ii}}}, \quad i=1, 2, \cdots, n \tag{3}$$

The calibration point of r max is $\sqrt{\frac{(n-p')F}{(n-p'-1)+F}}$ Where F is the F(1, n−p−1).

2.3.1 Externally Studentized Residuals

$$t_i = \frac{e_i}{\hat{\sigma}_{(i)}\sqrt{1-h_{ii}}} = r_i\sqrt{\frac{n-p'-1}{n-p'-r_i^2}}, \quad i=1, 2, \cdots, n. \tag{4}$$

Where $\hat{\sigma}_{(i)}{}^2 = \hat{\sigma}^2\left[\frac{n-p'-r_i^2}{n-p'-1}\right]$

The calibration point is $t(n-p'-1)$ or $\sqrt{F(1, n-p'-1)}$. t_i is a monotonic transformation of r_i.

3. DIAGNOSTIC MEASURES FOR DETECTING INFLUENTIAL CASES

3.1 Influence Measures Based On Remoteness of Points In The X−Y Space

3.1.1 Leverage : Diagonal Elements of Hat Matrix

$$h_{ii} = X^T{}_i(X^TX)^{-1}X_i, \quad i=1, 2, \cdots, n. \tag{5}$$

The calibration point is 0.2, $\frac{2p'}{n}$ or $\frac{n(p'-1)F+n-k}{n(p'-1)F+n(n-k)}$ Where F is the F(p'−1, n−p').

3.2.1 Mahalanobi's Distance

$$M_i = (n-2)\ (X_i-\overline{X})^T\ (X^TX)^{-1}(X_i-\overline{X})$$

$$\frac{n(n-2)}{n-1}\ \frac{h_{ii}-\frac{1}{n}}{1-h_{ii}}, \quad i=1, 2, \cdots, n. \tag{6}$$

Where $X^TX=\begin{bmatrix} \Sigma(X_{i1}-\overline{X}_1)^2 & \cdots & \Sigma(X_{i1}-\overline{X}_1)(X_{ip}-\overline{X}_p) \\ \vdots & & \vdots \\ \Sigma(X_{i1}-\overline{X}_1)(X_{ip}-\overline{X}_p) & \cdots & \Sigma(X_{ip}-\overline{X}_p)^2 \end{bmatrix}$

M_i is equivalent to h_{ii}

3.1.3 Modified Leverage

$$h_{ii}^* = z^T(Z^TZ)^{-1}z$$
$$= h_{ii} + \frac{e_i^2}{e^Te}, \quad i=1,\ 2,\ \cdots,\ n. \tag{7}$$

Where $Z=(X \vdots Y)$.

The calibration point is $\frac{2(p'+1)}{n}$

3.1.4 Weighted Squared Standardized Distance (Daniel et al., 1980)

$$WSSD_i = \sum_{j=1}^{n} \frac{\hat{\beta}_j^2(X_{ij}-\overline{X}_j)^2}{\dfrac{\sum_{j=1}^{n}(y_k-\bar{y})^2}{n-1}}, \ i=1,\ 2,\ \cdots,\ n. \tag{8}$$

In simple regression case, $WSSD_i$ is equivalent h_{ii}.

3.2 Influence Measures Based On The Sample Influence Curve

3.2.1 Cook's Distance (Cook, 1977)

$$D_i = \frac{(\hat{\beta}_{(i)}-\hat{\beta})^T\ (X^TX)\ (\hat{\beta}_{(I)}-\hat{\beta})}{p'\hat{\sigma}^2}$$
$$= \frac{(\hat{Y}_{(i)}-\hat{Y})^T\ (\hat{Y}_{(i)}-\hat{Y})}{p'\hat{\sigma}^2}$$
$$= \frac{1}{p'}\ \frac{h_{ii}}{1-h_{ii}}\ r_i^2, \quad i=1,\ 2,\ \cdots,\ n. \tag{9}$$

The calibration point is $F(p',\ n-p')$.

3.2.2 DEFITS (Belsley et al., 1980)

$$DEFITS_i = \frac{\hat{Y}_{(i)}-\hat{Y}}{\hat{\sigma}_{(i)}\sqrt{h_{ii}}}$$
$$= t_i\sqrt{\frac{h_{ii}}{1-h_{ii}}}, \quad i=1,\ 2,\ \cdots,\ n. \tag{10}$$

The calibration point is $2\sqrt{\frac{p'}{n}}$ or $2\sqrt{\frac{p'}{n-p'}}$.

$(DEFITS_i)^2 = \frac{(\hat{\beta}_{(i)}-\hat{\beta})^T(X^TX)(\hat{\beta}_{(I)}\hat{\beta})}{\hat{\sigma}_{(i)}^2}$. This measure differs from D_i by a scale factor and replacement of $\hat{\sigma}^2$ by $\hat{\sigma}_{(i)}^2$.

3.2.3 Modified DEFITS (Welsch, 1982)

$$DEFITS_i^* = DEFITS_i \sqrt{\frac{n-1}{1-h_{ii}}}, \quad i=1, 2, \cdots, n. \tag{11}$$

3.2.4 Modified Cook's Distance (Atkinson, 1981)

$$D_i^* = DEFITS_i \sqrt{\frac{n-p'}{p'}}, \ i=1, 2, \cdots, n. \tag{12}$$

The calibration point is $2\sqrt{\frac{n-p'}{p'}}$

In brief D_i measures influence of the ith case on $\hat{\sigma}$ only, whereas DEFITS, $DEFITS_i^*$, D_i^* measure the influence on $\hat{\beta}$ and $\hat{\sigma}$.

3.3 Influence Measures Based On Volume Of Confidence Ellipsoids

3.3.1 Andrews-Pregibon Statistic (Andrews et al., 1978)

$$AP_i = \frac{\det(Z_{(i)}^T Z_{(i)})}{\det(Z^T Z)}$$

$$= 1 - h_{ii}^*, \quad i=1, 2, \cdots, n. \tag{13}$$

AP_i is equivalent $(1-h_{ii}^*)$.

3.3.2 COVRATIO (Belsley et al., 1980)

$$COVRATIO_i = \frac{\det|\hat{\sigma}_{(i)}^2 (X_{(i)}^T X_{(i)})^{-1}|}{\det|\hat{\sigma}^2 (X^T X)^{-1}|}$$

$$= \left[\frac{n-p'-r_i^2}{n-p'-1}\right] \frac{1}{1-h_{ii}}, \quad i=1, 2, \cdots, n. \tag{14}$$

The calibration point is $|COVRATIO_i - 1| \geq \frac{3p'}{n-p'}$ or $|COVRATIO_i - 1| \geq \frac{3p'}{n}$

3.3.3 FVARATIO (Belsley et al., 1980)

$$FVARATIO_i = \frac{Var(\hat{Y}_{(i)})}{VarI(\hat{Y})}$$

$$= \frac{e_i^2}{\hat{\sigma}^2 t_i^2 (1-h_{ii})^2}, \ i=1, 2, \cdots, n. \tag{15}$$

The calibration point is $FVARATIO_i \leq 1 - \frac{3}{n}$ or $FVARATIO_i \geq 1 + \frac{2p+3}{n}$

3.3.4 Modified COVRATIO (Cook et al., 1980)

$$COVRATIO_i^* = \log \frac{\text{volume}\left[\frac{(\beta-\hat{\beta})^T (X^T X)(\beta-\hat{\beta})}{p'\sigma^2}\right]}{\text{volume}\left[\frac{(\beta-\hat{\beta}_{(i)})(X_{(i)}^T X_{(i)})(\beta-\hat{\beta}_{(i)})}{p'\sigma_{(i)}^2}\right]}$$

$$= -\frac{1}{2}\log(COVRATIO_i) + \frac{p'}{2}\log \frac{F(p', n-p')}{F(p', n-p'-1)}, \ i=1, 2, \cdots, n. \tag{16}$$

$COVRATIO_i^*$ is equivalent to $COVRATIO_i$

3.4 Influence Measures Based On The Likelihood Distance

3.4.1 Estimation of both β and σ^2 (Cook et al., 1982)

$LD_i(\beta, \sigma^2 0 = 2\{l(\tilde{\beta}, \tilde{\sigma}^2) - l(\tilde{\beta}_{(i)}, \tilde{\sigma}_{(i)}^2)\}$

$$= n\ln\frac{n(n-p'-r_i^2)}{(n-1)(n-p')} + \frac{(n-1)\ r_i^2}{(1-h_{ii})(n-p'-r_i^2)} - 1,\ i=1,\ 2,\ \cdots,\ n. \qquad (17)$$

Where $l(\tilde{\beta}, \tilde{\sigma}^2)$ is the log-likelihood evaluated at $\tilde{\beta}$ and $\tilde{\sigma}$ which are the maximum likelihood estimate of β and σ^2.

The calibration point is $X^2(p'+1)$.

3.4.2 Estimation of β.

$LD_i(\beta \mid \sigma^2) = 2\{l(\tilde{\beta}, \tilde{\sigma}^2) - \max\ l(\tilde{\beta}_{(i)}, \sigma^2\}$

$$= n\ln(1 + \frac{p'}{n-p'} D_i),\quad i=1,\ 2,\ \cdots,\ n. \qquad (18)$$

The calibration point is $X^2(p')$. $LD_i(\beta \mid \sigma^2)$ is equivalent to D_i.

4. DIAGNOSTIC HESURES FOR DETECTING OUTLIERS AMD INFLUENTIAL CASES

4.1 Proposed Measures Based On The PRESS

4.1.1 Squard Predicted Residuals (Squared PRESS Residuals)

$e_{(i)}^2 = [Y_i - X_i^T \tilde{\beta}_{(i)}]^2$

$$= \left[\frac{e_i^2}{1-h_{ii}}\right]^2,\ i=1,\ 2,\ \cdots,\ n. \qquad (19)$$

The PRESS measure is the sum of square of the PRESS residuals.

Allen (1974) used $PRESS = \Sigma e_{(i)}^2$ as a criterion for model selection, better models corresponding to relatively small values of PRESS.

We investigate the relationship between the ith squared PRESS residuals and the reduced Cook's distance.

$$e_{(i)}^2 = \frac{e_i^2}{1-h_{ii}}\ \frac{1}{1-h_{ii}} \qquad (20)$$

The reduced Cook's distance, except from the factor $p'\ \hat{\sigma}^2$, is

$$RD_i = \frac{e_i^2}{1-h_{ii}}\ \frac{h_{ii}}{1-h_{ii}} \qquad (21)$$

Knowing (20) and (21) the relationship of ith squared PRESS residuals and the reduced Cook's distance is very high.

By Belsley, Kuh, and Welsch (1980), $DEFIT_i$ and $DFBETA_i$ are related to PRESS residuals.

$$DEFIT_i = e_{(i)}\ h_{ii}$$

$$DFBETA_i = \frac{e_{(i)}(1-h_{ii})}{n-1}$$

So we propose the ith squared PRESS residuals as influence measure.

4.1.2 r_i^2

$$r_i^2 = \frac{e_{(i)}^2}{\frac{\hat{\sigma}^2}{(1-h_{ii})}},\quad i=1,\ 2,\ \cdots,\ n \qquad (22)$$

4.1.3 t_i^2

$$t_i^2 = \frac{e_{(i)}^2}{\frac{\sigma^2_{(i)}}{1-h_{ii}}}, \quad i=1,\ 2,\ \cdots,\ n \tag{23}$$

r_i^2 and t_i^2 are squared form of r_i and t_i, besides they are function of $e_{(i)}^2$. So we propose r_i^2 and t_i^2 as a diagnostic measure for detecting outliers and influential cases.

5. EXAMPLE

Montogomery and peck(1982) considered a set of data on the delivery times. The measured variables are Y=the time to service a vending machine;X_1=the number of items stocked by the machine;X_2=the distance traveled to reach it.

But we use modified delivery-time data to assume the linearity.

The data are reproduced in Table 1.

A linear model

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\epsilon$$

We now examine the several outlyingness and influence measures which we have described earlier. These results are show in Tables 2-7.

The results of analysis are as follows :

1. In Outlyingness measures based on residuals, t_i designates case 11 as a outlier.
2. In influence measures based on remo teness of points in the X−Y space, h_{ii}, M_i, h_{ii}^* and $WSSD_i$ clarify case 9 to be the most influential one.
3. *In influence measures based on the Influence curve,* $DEFITS_i$ and D_i^* declare cases 9 and 11 to influential ones, while $DEFITS_i*$ picks out only case 9. But D_i indicate that no case in influential.
4. In influence measures based on volume of confidence ellipsiods, AP_i, $COVRATIO_i$ and $FVARATIO_i$ declares cases 9 and 22 to be influential ones. But $COVRATIO_i^*$ indicate that no case is influential.
5. In influence measures based on the likelihood distance, No influential case is detected.
6. In proposed diagnostic measures, $e_{(i)}^2$, r_i^2 and t_i^2 indicate that case 11 is an outlier and influential one.

The effects of R^2 When deleting influential cases are shown in Table 8.

We conclude that case 11 is the most influential case.

t_i, $DEFITS_i$, D_i^* and proposed measures pickout case 11.

Table 1. Modified Delivery-Time Data : Montogomery et al. (1982)

CASE	X_1	X_2	Y
1	7	313600	16.68
2	3	48400	11.50
3	3	115600	12.03
4	4	6400	14.88
5	6	22500	13.75
6	7	108900	18.11
7	2	12100	8.00
8	7	44100	17.83
9	30	2131600	79.24
10	5	366025	21.50
11	16	473344	40.33
12	10	46225	21.00
13	4	65025	13.50
14	6	213444	19.75

15	9	200704	24.00
16	10	602176	29.00
17	6	40000	15.35
18	7	17424	19.00
19	3	1296	9.50
20	17	592900	35.10
21	10	19600	17.90
22	26	656100	52.32
23	9	202500	18.75
24	8	403225	19.83
25	4	22500	10.75

Table 2. Outlyingness Measure Based On Residuals

CASE	e_i	r_i	t_i	$\hat{\sigma}_{(i)}^2$
1	−3.86	−1.69	−1.77	5.03
2	0.39	0.17	0.17	5.78
3	−0.04	0.02	−0.02	5.79
4	2.95	1.29	1.32	5.35
5	−1.23	−0.54	−0.53	5.71
6	0.49	0.21	0.21	5.77
7	−1.19	−0.53	−0.52	5.71
8	1.13	0.50	0.49	5.72
9	0.35	0.35	0.34	5.75
10	3.04	1.38	1.41	5.29
11	4.82	2.18	2.41*	5.52
12	0.04	0.02	0.02	5.79
13	0.74	0.33	0.32	5.76
14	2.05	0.90	0.90	5.57
15	2.25	0.98	0.98	5.53
16	0.12	0.05	0.05	5.79
17	0.12	1.18	0.05	5.79
18	2.68	−0.42	1.19	5.78
19	−0.94	−1.59	−0.41	5.74
20	−3.52	−1.21	−1.65	5.12
21	−2.68	0.06	−1.23	5.40
22	0.10	−1.32	0.06	5.79
23	−3.02	−1.49	−1.34	5.33
24	−3.39	−0.62	−1.54	5.20
25	−1.41		−0.61	5.68

Table 3. Influence Measure on Remoteness of Points in the X−Y Space

CASE	h_{ii}	M_i	h_{ii}^*	$WSSD_i$
1	0.057	0.43	0.18	0.027
2	0.075	0.91	0.08	0.313
3	0.087	1.23	0.09	0.293
4	0.060	0.51	0.13	0.243
5	0.054	0.35	0.07	0.112
6	0.046	0.15	0.05	0.046

7	0.088	1.26	0.10	0.433
8	0.055	0.38	0.07	0.067
9	0.813*	99.03*	0.81*	6.543*
10	0.117	2.09	0.19	0.124
11	0.111	1.91	0.30	0.466
12	0.102	1.65	0.10	0.053
13	0.062	0.56	0.07	0.221
14	0.053	0.33	0.09	0.065
15	0.045	0.13	0.09	0.004
16	0.093	1.40	0.09	0.103
17	0.051	0.28	0.05	0.106
18	0.061	0.54	0.12	0.077
19	0.070	0.77	0.08	0.332
20	0.110	1.88	0.21	0.645
21	0.113	1.97	0.17	0.063
22	0.511*	23.08	0.51	2.574
23	0.045	0.13	0.12	0.001
24	0.063	0.59	0.16	0.019
25	0.060	0.51	0.08	0.236

Table 4. Influence Measure Based on the Influence Curve

CASE	D_i	$DEFITS_i$	$DEFITS_i^*$	D_i^*
1	0.057	−0.435	−2.19	−1.18
2	0.001	0.048	0.25	0.13
3	0.000	−0.006	−0.03	−0.02
4	0.036	0.334	0.17	0.91
5	0.005	−0.126	−0.64	−0.34
6	0.001	0.046	0.23	0.14
7	0.009	−0.162	−0.83	−0.44
8	0.005	0.118	0.60	0.32
9	0.174	0.708*	8.02*	1.92*
10	0.084	0.513	2.68	1.39
11	0.197	0.849*	4.41	2.30*
12	0.000	0.006	0.03	0.02
13	0.002	0.082	0.42	0.22
14	0.015	0.212	1.07	0.58
15	0.015	0.213	1.07	0.58
16	0.000	0.017	0.09	0.05
17	0.000	0.012	0.06	0.03
18	0.030	0.303	1.53	0.82
19	0.004	−0.113	−0.57	−0.31
20	0.104	−0.581	−3.02	−1.58
21	0.063	−0.439	−2.29	−1.19
22	0.001	0.059	0.41	0.16
23	0.027	−0.290	−1.45	−0.79
24	0.050	−0.398	−2.01	−1.08
25	0.006	−0.154	−0.08	−0.42

Table 5. Influence Measures Based on Volume of Confidence Ellipsoids

CASE	AP_i	$COVRATIO_i$	$FVARATIO_i$	$COVRAIO_i$
1	0.82	0.803	0.966	0.043
2	0.92	1.238	1.132	−0.0046
3	0.91	1.259	1.148	−0.0546
4	0.87	0.966	1.031	0.0029
5	0.93	1.168	1.093	−0.0383
6	0.95	1.198	1.095	−0.0438
7	0.90	1.213	1.134	−0.0465
8	0.93	1.176	1.096	−0.0398
9	0.19*	6.046*	5.568*	−0.3950
10	0.81	0.992	1.085	0.0029
11	0.70	0.623	1.124	−0.0982
12	0.90	1.280	1.168	−0.0582
13	0.93	1.066	1.112	−0.0185
14	0.91	1.085	1.065	−0.0223
15	0.91	1.053	1.049	−0.0158
16	0.91	1.267	1.156	−0.0560
17	0.95	1.211	1.105	−0.0462
18	0.88	1.006	1.115	−0.0059
19	0.92	1.207	1.118	−0.0455
20	0.79	0.896	1.042	0.0192
21	0.83	1.102	1.104	−0.0257
22	0.49	2.350*	2.144*	−0.1900
23	0.88	0.940	1.011	0.0088
24	0.84	0.892	1.005	0.0202
25	0.92	1.160	1.094	−0.0368

Table 6. Influence Measures Based on the Likelihood Distance

CASE	$LD_i(\beta\ \sigma^2)$	$LD_i(\beta \mid \sigma^2)$
1	0.340	0.1940
2	0.022	0.0034
3	0.021	0.0000
4	0.143	0.1220
5	0.028	0.0170
6	0.021	0.0030
7	0.039	0.0310
8	0.027	0.0170
9	0.600	0.5860
10	0.333	0.2850
11	1.366	0.6630
12	0.021	0.0000
13	0.023	0.0070
14	0.052	0.0510
15	0.052	0.0510
16	0.021	0.0003
17	0.021	0.0000
18	0.113	0.1020

19	0.028	0.0140
20	0.470	0.3520
21	0.228	0.2140
22	0.024	0.0030
23	0.120	0.0920
24	0.236	0.1700
25	0.034	0.0270

Table 7. Proposed Diagnostic Measures

CASE	$e_{(i)}^2$	r_i^2	t_i^2
1	16.755	2.86101	3.14890
2	0.178	0.02981	0.02859
3	0.002	0.00033	0.00031
4	9.849	1.67644	1.73500
5	1.691	0.28967	0.28005
6	0.264	0.04561	0.04373
7	1.703	0.28124	0.27190
8	1.430	0.24470	0.23658
9	3.503	0.11862	0.11373
10	11.853	1.89519	1.97849
11	29.396*	4.73213*	4.73213*
12	0.002	0.00033	0.00031
13	0.622	0.10565	0.10129
14	4.686	0.80356	0.79676
15	5.551	0.95993	0.95993
16	0.018	0.00296	0.00281
17	0.016	0.00275	0.00261
18	8.146	1.38508	1.32797
19	1.022	0.17211	0.16640
20	15.643	2.52101	2.72580
21	9.150	1.46800	1.50621
22	0.042	0.00372	0.00354
23	10.067	1.74087	1.80169
24	13.089	2.22081	2.35927
25	2.250	0.38298	0.37339

Table 8. Effects of R^2 When Deleting Influential Case

Case deleted	R^2
None	97.9%
9, 11, 22	90.0%
9, 22	91.0%
9	95.0%
9, 11	95.4%
22	97.5%
11, 22	98.0%
11	98.3%

6. CONCULUSION

We reviewed the various diagnostic measures for detecting outliers and influential cases in linear regression. The various measures didn't pickout same influential cases. So In this paper, we showed the effects of R^2 When deleting influential cases. We recommend three proposed measures for detecting outliers and influential cases.

References

Allen, D. M.(1974), "The Relationship Between Variable Selection and Data Augumentation and a Method for Prediction," *Technometrics,* 16, 125-27.

Andrew, D. F., and Pregibon, D.(1978), "Finding Ouliers That Matter," *Journal of the Royal Statistical Society,* (B), 40, 85-93.

Atkinson, A. C.(1981), "Two Graphical Displays for Outlying and Influential Observations in Regression," *Biomerika,* 68, 13-20.

Belsley, D. A., Kuh, E., and Welsch, R. E.(1980), *Regression Diagnostics : Indentifying Influenial Data and Sources of Collinearity,* New York : John Wiley & Sons.

Cook, R. D.(1977), "Detection of Influential Observations in Linear Regression," *Technometrics, 19,* 15-18.

______, and Weisberg, S.(1980), "Characterization of an Empirical Influence Function for Detecting Influential cases in Regression," *Technometrics,* 22, 495-508.

______________(1982), *Residuals and influence in Regression,* London : Chapman and Hall.

Daniel, C., and Wood, F. S.(1980), *Fitting Equations to Data,* 2nd ed., New York : John Weiley & Sons.

Montogomery, D. C., and Peck, E. A.(1982), *Introduction to Linear Regression Aalysis,* New York : Jhno Wiley & Sons.

Welsch, R. E.(1982), *Influence Functions and Regression Diagnostics in Modern Data Analysis,* New York : Academic Press.