Journal of the KSQC
Vol. 16, No.2 1988

A Study on Detection of Outliers and Influential Observations in Linear Models*

Eun M. Kang**
Sung H. Park***

ABSTRACT

A new diagnostic statistic for detecting outliers and influential observations in linear models is suggested and studied in this paper. The proposed statistic is a weighted sum of two measures ; one is for detecting outliers and the other is for detecting influential ovservations. The merit of this statistic is that it is possible to distinguish outliers from influential observations. This statistic can be used for not only regression models but also factorial design models. A Monte Carlo simulation study is reported to suggest critical values for detecting outliers and influential observations for simple regression models when the number of observations is 11, 21, 31, 41 or 51.

1. Introduction for Diagnostic Measures

Recently a great number of research papers have been published on the area of outliers and influential observations for diagnostic purposes, and there still remain many unsolved problems. It is known that observations of, in the opinion of the investgator, standing apart from the bulk of the data have been called "outliers". It is also known that observations are judged as

* This work was partially supported by the Ministry of Education, through the Research Institute for Basic Sciences, Seoul National University, 1987-1988.
** Department of Statistics, Sung Shin Women's University.
*** Department of Computer Science and Statistics, Seoul National University.

"influential" if important features of the analysis are substantially altered when the observations are deleted. A great deal of measures have been proposed to detect outliers and influential observations for regression models and/or factorial experiments.

Suppose the linear regression model can be written as

$$\underline{y}=X\underline{\beta}+\underline{\varepsilon} \tag{1.1}$$

where $\underline{y}$ is an $n\times 1$ vector of observations, X is an $n\times p$ full rank matrix of known constants, $\underline{\beta}$ is a p vector regression coefficients and $\underline{\varepsilon}$ is an $n\times 1$ vector of randomly distributed errors such that $E(\underline{\varepsilon})=\underline{0}$ and $V(\underline{\varepsilon})=I\sigma^2$. In fitting the model (1.1) by least squares, we ususally obtain the fitted or predicted value from $\underline{\hat{y}}=X\underline{\hat{\beta}}$ where $\underline{\hat{\beta}}=(X'X)^{-1}X'\underline{y}$. From this, it is simple to see that

$$\underline{\hat{y}}=H\underline{y} \tag{1.2}$$

$$\underline{e}=\underline{y}-\underline{\hat{y}}=(I-H)\underline{y} \tag{1.3}$$

where $H=X(X'X)^{-1}X'$ is the hat matrix and $\underline{e}$ is the residual vector.

Note that $\underline{\hat{y}}$ is the perpendicular projection of $\underline{y}$ into the subspace generated by columns of X. Since, H is symmetric and idempotent, we can write

$$h_{ii}=h_{ii}^2+\sum_{j\neq i}h_{ij}^2 \tag{1.4}$$

and it is clear that $0\leq h_{ii}\leq 1$. In this case, $\mathrm{rank}(H)=\mathrm{rank}(X)=p$ and hence, $\mathrm{trace}(H)=p$. The average of diagonal elements h_{ii} of the hat matrix, then, is p/n. Experience suggests that a reasonable rule of thumb for large h_{ii} is $h_{ii}>2p/n$. Thus we determine a high leverage point by looking at the diagonal elements of H and paying particular attention to any design point for which $h_{ii}>2p/n$. We may say that if h_{ii} is large, the data point may be considered as influential.

For a measure of an outlier, we often use the standardized residuals

$$r_i=\frac{e_i}{s\sqrt{1-h_{ii}}} \tag{1.5}$$

where $s=\sqrt{\underline{e}'\underline{e}/(n-p)}$. However, s^2 tends to overestimate σ^2 when there exists an outlier. For such case, $s_{(i)}^2$ is a better choice as an estimate of σ^2, were $s_{(i)}^2$ is the residual mean square error of $n-1$ observations after discarding the ith possible outlier case. Then we obtain the studentized residual

$$r_i^*=\frac{e_i}{s_{(i)}\sqrt{1-h_{ii}}} \tag{1.6}$$

which is t-distributed with $(n-p-1)$ degrees of freedom.

A useful measure of influence which is called the Cook's statistic

$$D_I=(\underline{\hat{\beta}}-\underline{\hat{\beta}}_{(I)})'X'X(\underline{\hat{\beta}}-\underline{\hat{\beta}}_{(I)})/ps^2 \tag{1.7}$$

is obtained by Cook(1977, 1979), where $\underline{\hat{\beta}}_{(I)}$ is the least squares estimate of $\underline{\beta}$ obtained by deleting k rows and k observations indexed by I from X and $\underline{y}$, respectively, and s^2 is the least squares estimate of σ^2 in the full model. If there is a single observation deleted, D_I is written as D_i. Cook suggests that if the observed D_i is equal to or greater that $F(p,\ n-1\ ;\ \alpha)$ where α is less than 0.5, then y_i may be significant as an influential observation.

Andrews and Pregibon(1978) suggest a statistic using the ratio

$$R_I=\frac{(n-p-k)s^2_{(I)}|X'_{(I)}X_{(I)}|}{(n-p)s^2|X'X|} \tag{1.8}$$

for identifying subsets of k influential cases. The rationale for this measure is based on the idea that the deletion of an outlier in $\underline{y}$ will result in a marked reduction in the residual sum of squares and the deletion of a remote point in X will produce a similar change in $|X'X|$. This quantity is dimensionless. Geometrically $1-(R_I)^{\frac{1}{2}}$ corresponds to the proportion of the volume generated by $(X\ :\ \underline{y})$ attributable to the indexed k observations. Accordingly, small values of R_I are associated with deviants and/or influential observations.

For multiple outlier case Gentleman and Wilk (1975 b) suggest Q_k

$$Q_k=RSSc-RSSm \tag{1.9}$$

where k indicates k outliers, RSSc is the residual sum of squares when the complete set of original data is used to fit the specified model, and RSSm is the residual sum of squares when the extreme observation are regarded as missing.

We write the basic model

$$E(\underline{y})=E\begin{bmatrix}\underline{y}_1\\ \underline{y}_2\end{bmatrix}=\begin{bmatrix}X_1\\ X_2\end{bmatrix}\beta \tag{1.10}$$

where X_1 is an $(n-k)\times p$ matrix which contains no outliers, and X_2 is a $k\times p$ matrix which contains k outliers. Then we can express Q_k as

$$\begin{aligned}Q_k&=Q_{k_1}+Q_{k_2}\\ &=\underline{e}_2'\underline{e}_2+\underline{e}_1'X_1(X_1'X_1)^{-1}X_1'\underline{e}_1\end{aligned} \tag{1.11}$$

where $Q_{k_1}=\underline{e}_2'\ \underline{e}_2$, $Q_{k_2}=\underline{e}_1'X_1(X_1'X_1)^{-1}X_1\underline{e}_1$, and

$$\underline{e}=\begin{bmatrix}\underline{e}_1\\ \underline{e}_2\end{bmatrix}=(I-R)\underline{y}$$

$$= \begin{bmatrix} I-R_{11} & -R_{12} \\ -R_{21} & I-R_{22} \end{bmatrix} \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \end{bmatrix} \tag{1.12}$$

Here, $R = X(X'X)^{-1}X'$ and $R_{ij} = X_i(X'X)^{-1}X_j'$. It is not difficult to show that

$$Q_{k2} = (\underline{\tilde{\beta}} - \underline{\hat{\beta}})' X'X (\underline{\tilde{\beta}} - \underline{\hat{\beta}}) \tag{1.13}$$

where $\tilde{\beta} = (X_1'X_1)^{-1}X_1'\underline{y}_1$. Note that if Q_{k1} is large, the k observations may be outliers, and if Q_{k2} is large, the k observations may be influential observations.

2. Proposition of a New Statistic

For regression models, the standardized residual r_i and the studentized residual r_i^* only serve to detect outliers, while such measures as h_{ii}, the Cook's statistic D_I are only used for detecting influential observations. Such measures as the Andrews-Pregibon's R_I and the Gentleman and Wilk's Q_k may detect the observations which are outliers and/or influential observations. However, in practice we want to know which observations are outliers and which observations are influential. Statistics which can distinguish outliers from influential observations have not been suggested.

As mentioned in the previous section Q_k is decomposed into Q_{k1} and Q_{k2}, where Q_{k1} mainly detects outliers and Q_{k2} mainly detects influential observations. However, the magnititude of Q_{k1} and Q_{k2} heavily depends on the unit of observations. Hence, to make Q_{k1} and Q_{k2} scale invariant, we need to divide them by some scaling factor. We propose the following statistic which is a weighted sum of Q_{k1} and Q_{k2} divided by some scaling factor

$$WQ_k = wQ_{k1}/(s.f.) + (1-w)Q_{k2}/(s.f.) \tag{2.1}$$

where w is the weight factor, *i.e.* $0 \le w \le 1$ and $s.f.$ is a scaling factor.

Now we choose an appropriate scaling factor for detecting outliers and influential observations. The appropriate scaling factor we want to propose is $ks_{(I)}^2$ where $s_{(I)}^2$ is the residual sum of squares of the reduced model which does not include the observations indexed by I.

Note that when $w=0$, WQ_k *becomes* $Q_{k2}/ks_{(I)}^2$ which is similar to the cook's D_I in the equation (1.2). When $w=1$, WQ_k becomes $Q_{k1}/ks_{(I)}^2$ which is the sum of squares of the largest k residuals divided by $ks_{(I)}^2$. Hence, $Q_{k1}/ks_{(I)}^2$ can detect k outliers. When $w=0.5$, WQ_k becomes $Q_k/2ks_{(I)}^2$ which behaves like the statistic Q_k(it detects the same points as Q_k). However, the most important point of this statistic is that as the weight changes from 0 to 1, it can show the influential observations at first and then gradually changes to outliers. Therefore, we can easily distinguish outliers from influential observations.

Here we are interested in the probability distribution of the maximum value of WQ_k, for convenience, we may write this as max WQ_k. However, it is difficult to obtain the exact distribution of max WQ_k. Hence, in this paper we want to show the empirical probability distribution of max WQ_k by Monte Carlo simulation, and we will provide the critical values for

some given significance levels which can be used for hypothesis testing of outliers.

Next, it is of interest to compare the diagnostic measures with the proposed statistic. The value of Q_k consists of two parts, the outlier part and the influential part. However, since the value of Q_k seems to be dominated by the outlier part, Q_k is categorized as a measure of detecting outliers in Table 1. According to Table 1 each of all the diagnostic statistics is some function of r_i, r_i^* and h_{ii}. And WQ_k can be represented in a similar form. When $w=0$, WQ_k is $h_{ii}r_i^{*2}$, when $w=0.5$, WQ_k becomes $r_i^{*2}/2$ and when $w=1$, $WQ_k=(1-h_{ii})r_i^{*2}$.
Table 2 shows the relationships among D_I, R_I, Q_k and WQ_k, when the number of outliers or infulential observations are greater than 1.

Table 1. Comparisons of Measures for Detecting an Outlier and/or an influential observation

Mesures for detection	Measures for detection of an influential observation	Proposed Statistic
1. $r_i=\frac{e_i}{s\sqrt{1-h_{ii}}}$ 2. $r_i^*=\frac{e_i}{s_{(i)}\sqrt{1-h_{ii}}}$ 3. $Q_k=s^2r_i^2=s_{(i)}^2r_i^{*2}$	1. h_{ii}=leverage 2. $D_i=\frac{h_{ii}}{p(1-h_{ii})}r_i^2$ 3. $R_i=(1-h_{ii})(1-\frac{r_i^2}{n-p})$	1. When $w=0$, $WQ_k=h_{ii}r_i^{*2}$ 2. When $w=0.5$ $WQ_k=r_i^{*2}/2$ 3. When $w=1.0$ $WQ_k=(1-h_{ii})r_i^{*2}$

Table 2. Comparisons of Measures for more than 1 outlier and/or influential observation

Mesures for detection of outliers	Measures for detection of influential observation	Proposed Statistic
1. $Q_k=\underline{e}_2'(I-R_{22})^{-1}\underline{e}_2$ $=\underline{e}_2'\underline{e}_2+\underline{e}_1'X_{(I)}(X_{(I)}'X_{(I)})^{-1}X_{(I)}'\underline{e}_1$	1. $D_I=\frac{(\hat{\underline{\beta}}-\hat{\underline{\beta}}_{(I)})'X'X(\hat{\underline{\beta}}-\hat{\underline{\beta}}_{(I)})}{ps^2}$ 2. $R_I=\frac{\lvert X_{(I)}'X_{(I)}\rvert}{\lvert X^{*\prime}X^*\rvert}$ $=(1-\frac{Q_k}{RSS})\lvert I-R_{22}\rvert$	1. When $w=0.0$ $WQ=\frac{(\hat{\underline{\beta}}-\hat{\underline{\beta}}_{(I)})'X_{(I)}'X_{(I)}(\hat{\underline{\beta}}-\hat{\underline{\beta}}_{(I)})}{ks_{(I)}^2}$ $=\frac{\underline{e}_1'(X_{(I)}'X_{(I)})^{-1}X_{(I)}'\underline{e}_1}{ks_{(I)}^2}$ 2. When $w=0.5$ $WQ_k=\frac{Q_k}{2ks_{(I)}^2}$ 3. When $w=1.0$ $WQ_k=\frac{\underline{e}_2'\underline{e}_2}{ks_{(I)}^2}$

3. Example

The statistic max WQ_k will be obtained from the analysis of a set of 21 observations(x, y) which is similer to the data set given by Mickey, Dunn and Clark (1967). The observations appear in Table 3 and are plotted in Figure 1. A straight line regression model is fitted to the full set of data and then to the 20 data points remained when each observation is deleted in turn. Our test statistic max WQ_k is obtained where the scale factor(s. f.) is $ks^2_{(i)}$ and k is the assmed number of outliers. Table 4 shows the weights ranged from 0 to 1 and max WQ_k, and the deleted observation number and its corresponding significant probability. When the weights are small $(0 \leq w < 0.3)$, the number 18 is deleted. However, when the weights become larger $(w > 0.3)$, the deleted number changes from 18 to 19. The reason for this is that the residual for observation 19 is too large than any others.

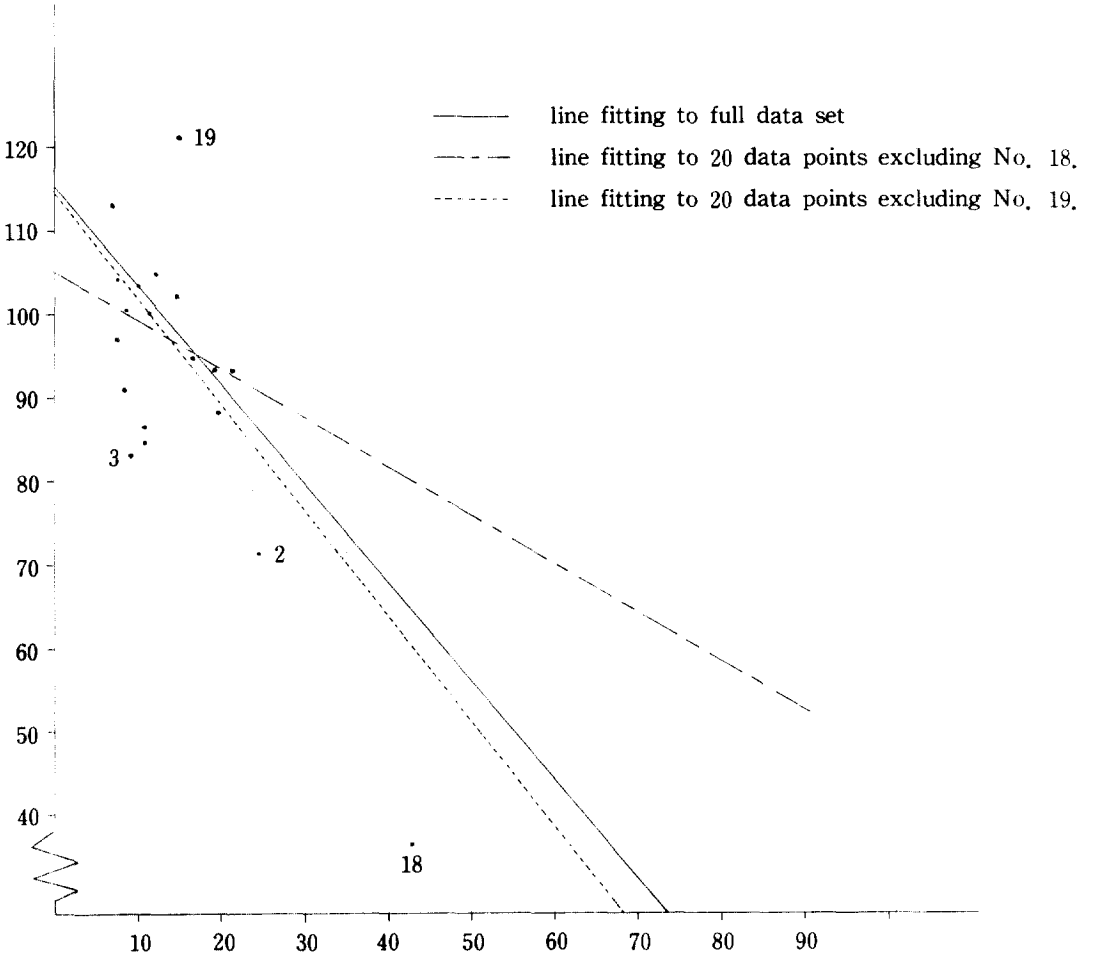

Figure 1. Plot of Example Data

For the removal of two cases, the results are summarized in Table 5. The results show that the deleted ovservations are varying with respect to w. It seems that the (2, 18) observations are most influential and (3, 19) are outliers. Here we also note that the significant probabilities are not small, therefore in this procedure the type I error may be greater than the often used values of 0.05, and 0.01. Based on the results of this example, we suggest that, to detect

Table 3. Age at First Work(x) and Gesell Adaptive Score(y)

Case	x	y	Case	x	y
1	15	95	11	7	113
2	26	71	12	9	96
3	10	83	13	10	83
4	9	91	14	11	84
5	15	102	15	11	102
6	20	87	16	10	100
7	18	93	17	12	105
8	11	100	18	42	35
9	8	104	19	17	121
10	20	94	20	11	86
			21	10	100

Table 4. Detected Observation and Significant Probability when one point Detection for Example Data

given weight	max. WQ_k	deleted observation	significant probability
0.0	2.609	18	0.07
0.1	2.480	18	0.09
0.2	2.942	19	0.08
0.3	4.75	19	0.04
0.4	5.208	19	0.06
0.5	6.341	19	0.04
0.6	7.474	19	0.04
0.7	8.607	19	0.04
0.8	9.740	19	0.03
0.9	10.873	19	0.03
1.0	12.005	19	0.04

the outliers and influential observations, the significance level α may be set to some big values such as $\alpha=0.1$ or 0.2 or even greater values. Table 6 shows the detected points when other statistics are used in this example.

Table 5. Detected Observations and Significant Probability when Two Point Detection for Example

given weight	max. WQ_k	deleted observation	significant probability
0.0	3.108	2, 18	0.04
0.1	2.881	2, 18	0.06
0.2	2.653	2, 18	0.14
0.3	3.012	18, 19	0.18
0.4	3.694	18, 19	0.20
0.5	4.376	18, 19	0.19
0.6	5.161	3, 19	0.19
0.7	5.986	3, 19	0.17
0.8	6.810	3, 19	0.14
0.9	7.635	3, 19	0.15
1.0	8.459	3, 19	0.17

Table 6. Detected Observations for Example in Other Statistics (the values of test statistics for detected points are given within the parentheses)

Statistics	Detected	Points
Q_K	19	(12.68)
	18. 19	(8.75)
R_i	18	(0.27)
	2. 18	(0.07)
D_I	18	(3.268)
	2. 18	(11.29)
r_i	19	(2.79)
r_i^*	19	(3.55)
h_{ii}	18	(0.65)

In one point detection case, the detected ovservation is the number 19 when using the outlier detecting statistic Q_k, r_i and r_i^*, while the number 18 is detected by using the statistics of influential observations such as D_I, R_I, and the leverage h_{ii}. From these results, the number 19 is the most outlying case and the number 18 is the most influential and remote point which has been already shown in Table 4.

For two points detection R_I and D_I detect the numbers 2 and 18 where Q_k detects the numbers 18 and 19. In Table 5 the results include these points and in addition, when $w>0.5$, the points 3 and 19 are detected.

4. Probability Distribution of the Proposed Statistic by Monte Carlo Simulation

There are so many possible cases we have to consider for simulation such as simple regression, multiple regression, factorial experiments, and so on. However, we only simulate the simple regression case in this paper.

We simulate the simple regression case when the design points are given in Table 3. The values of the dependent variable $\underline{y}$ are generated with the intercept 20 and the slope 2.0 using the normal random variate generator GGNML in IMSL subroutine. The simulated critical values for the given data set are given in Table 7(A). And the empirical probability density function of WQ_k is plotted in Figure 2. From Figure 2 the distribution of WQ_k is skewed to the right when the weight w is small and it moves to the right hand side when w becomes larger and it tends to be symmetric.

We have obtained these results by simulation runs, employing the following procedure.

(1) The 21 values of $\underline{y}$ are generated using the intercept 20 and the slope 2.0 where the error terms are generated by the normal random variate generator with standard deviation 5.0.

(2) For one point detection the reduced residual sum of squares is calculated from $Q_k = \underline{e}_2' (I-R_{22})^{-1}\underline{e}_2$. In one point detection case, it is easily obtained using the leverage h_{ii} and the residual of the detected point in the full model e_i, *i.e.* $e_i^2/(1-h_{ii})$. For two points detection, the reduced residual sums of squares can be obtained by the partition of hat matrix H.

(3) Then, from the reduced residual sum of squares, the WQ_k is obtained by using the scaling factor $ks_{(I)}^2$.

(4) For the generation of a set of data y, the maximum WQ_k is obtained from $\binom{n}{k}$ combinations.

(5) The simulation is repeated until 1,000 values of WQ_k are obtained, and percentage points are then found.

Next, the design sets B, C, and D in Figure 3 are given to compare these critical values, whether they are design-dependent or not. The data sets B, C, and D are various dispersion types of independent variables as shown in Fugure 3. The dispersion of A is skewed to the right and that of D is skewed to the left where B and C are evenly distributed. From Table 7, we can find that the critical values are almost equal when $w \leq 0.2$. Hence, we know from these results that the distribution of WQ_k is invariant to the type of dispersion of the design points.

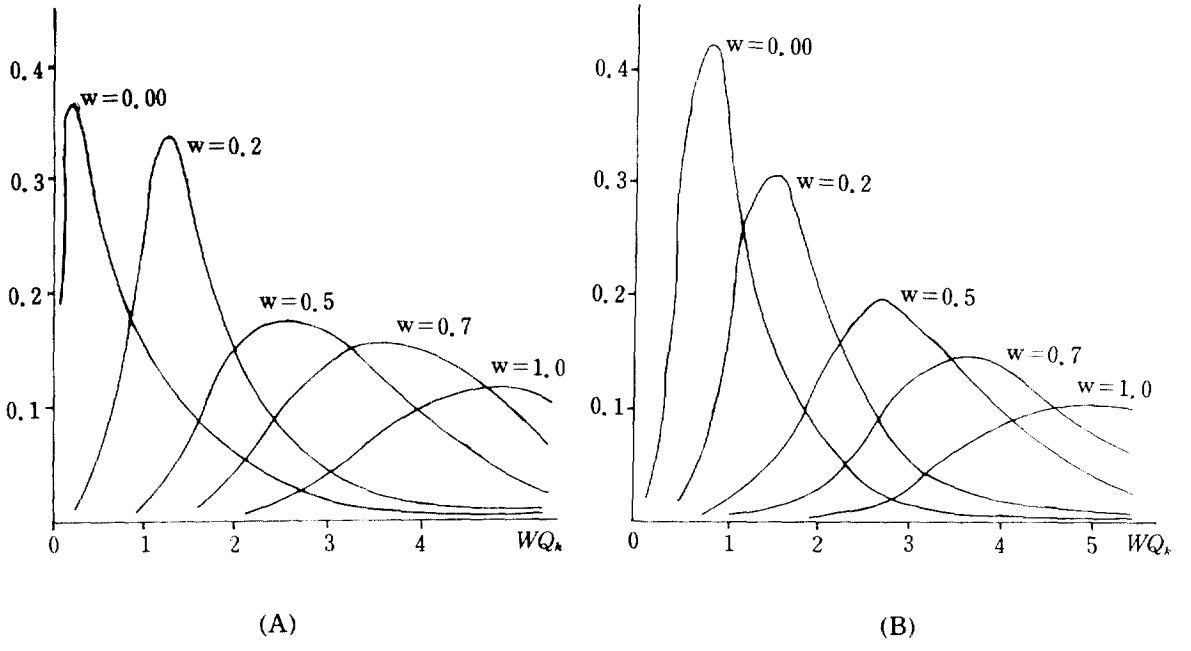

Figure 2. Empirical Probability Distribution of WQ_k with Various Weights in Simple Regression When the Number of Data Points is 21. (A) for One point Detection (B) for Two Points Detection

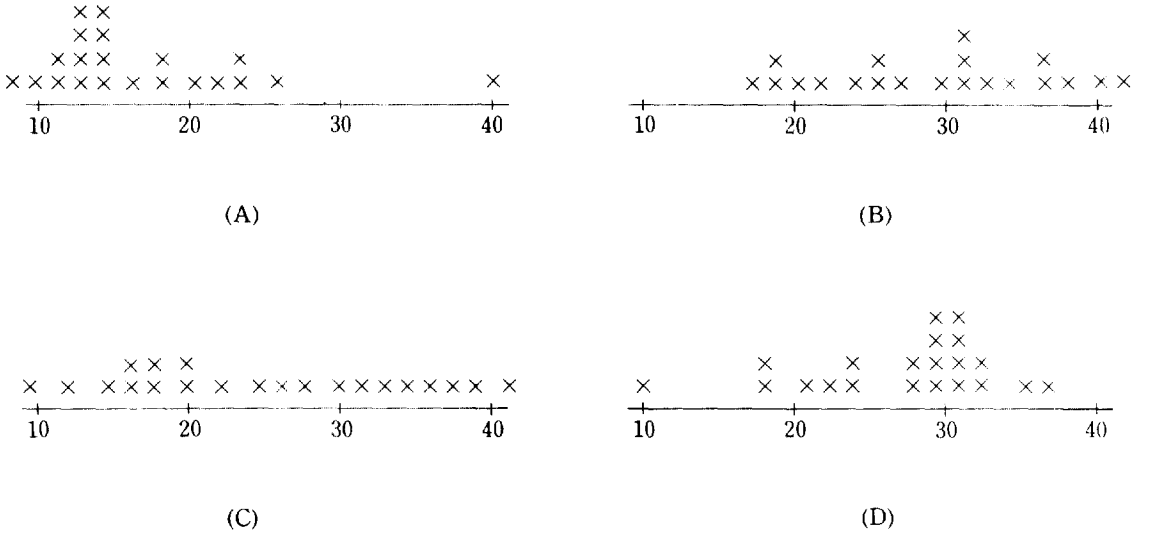

Figure 3. Scatter Plot of the Data Sets (A) – (D) When the Number of Data Points 21.

Next we want to show whether the number of design points has impact on the critical points or not. The randomly chosen design points for the number of points 11, 31, 41 and 51 are investigated. We obtain the critical values using the same routine from (1) through (5). Table 8 shows the critical values for the data sets with the number of points 11 to 51. Especially, in the case of 21, the tabulated values are the average of four critical values in Table 7. It is observed that for small w the critical values are in some degree shifted to smaller values when the number of points increases. On the other hand, for large w, the critical values are somewhat shifted to larger values when the number of points decreases. Hence we note that the distribution of max WQ_k under the null hypothesis(H_0 : There is no outlier), has the larger variance when the number of points are smaller.

In two points detection, the critical values for the data sets (A) and (C) are given in Table 9. The values are almost equal except when $w=0.0$ and 0.1. Table 10 gives the critical values for two points detection in the data sets with the number of points 11 to 51 and it show the similar pattern to one point case.

Tavle 7. The Critical Values of the Data Sets A, B, C, D for One Point Detection

data set	$\alpha \backslash w$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
A	0.05	3.02	3.32	3.54	4.36	5.42	6.17	7.40	8.26	8.90	10.3	11.7
	0.10	2.14	2.30	2.81	3.68	4.48	5.15	6.11	7.07	7.51	8.62	9.82
	0.30	0.91	1.39	1.91	2.51	3.03	3.69	4.16	4.86	5.27	5.84	6.94
	0.50	0.56	1.07	1.55	1.96	2.45	2.95	3.37	3.91	4.24	4.80	5.51
B	0.05	2.05	2.67	3.50	4.44	5.72	6.38	7.03	7.86	9.48	10.1	11.9
	0.10	1.53	2.08	2.78	3.59	4.53	5.04	5.88	6.83	7.63	8.49	9.82
	0.30	0.84	1.36	1.86	2.51	3.07	3.64	4.25	4.88	5.34	5.93	6.83
	0.50	0.59	1.08	1.51	2.00	2.49	2.91	3.44	3.95	4.31	4.82	5.46
C	0.05	1.49	2.29	3.29	4.44	5.25	6.29	7.15	8.19	8.89	10.4	12.0
	0.10	1.18	1.87	2.62	3.51	4.55	5.25	5.88	7.07	7.29	8.39	9.87
	0.30	0.74	1.31	1.90	2.47	3.14	3.18	4.18	4.85	5.23	5.86	6.73
	0.50	0.57	1.06	1.53	2.00	2.47	2.90	3.43	3.91	4.29	4.77	5.44
D	0.05	2.54	2.89	3.38	4.30	5.37	6.43	7.50	8.33	8.74	10.4	12.2
	0.10	1.73	2.20	2.84	3.64	4.29	5.21	5.85	6.78	7.45	8.29	10.2
	0.30	0.87	1.36	1.88	2.46	3.04	3.60	4.24	4.86	5.44	5.78	6.68
	0.50	0.58	1.06	1.50	1.96	2.44	2.90	3.42	3.93	4.35	4.80	5.44

Table 8. Critical Values for One Point Detection for Number of Data Points 11 to 51.

number of point	$\alpha \backslash w$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
11	0.05	2.87	3.88	4.64	5.97	7.03	8.08	8.69	9.20	10.70	11.16	13.70
	0.10	2.39	2.94	3.62	4.31	5.14	6.08	6.58	7.28	8.06	9.06	9.99
	0.30	1.32	1.77	2.15	2.68	3.17	3.76	4.03	4.41	4.87	5.69	5.99
	0.50	0.97	1.32	1.66	2.08	2.35	2.80	3.00	3.29	3.71	4.20	4.47
21	0.05	2.28	2.79	3.43	4.39	5.44	6.32	7.27	8.16	9.00	10.3	11.95
	0.10	1.65	2.11	2.76	3.61	4.46	5.16	5.93	6.94	7.47	8.45	9.93
	0.30	0.84	1.36	1.89	2.49	3.07	3.53	4.21	4.86	5.32	4.85	6.80
	0.50	0.58	1.07	1.52	1.98	2.46	2.92	3.42	3.93	4.30	4.80	5.46
31	0.05	1.08	1.84	2.86	4.12	4.79	6.20	7.25	8.10	9.39	10.06	11.383
	0.10	0.84	1.58	2.47	3.35	4.09	5.24	6.26	7.08	7.99	8.63	9.76
	0.30	0.57	1.17	1.87	2.49	3.09	3.85	4.52	5.23	5.75	6.49	7.10
	0.50	0.46	0.95	1.41	2.04	2.55	3.16	3.75	4.35	4.73	5.28	5.88
41	0.05	0.93	1.71	2.76	3.76	5.05	6.09	7.28	8.62	9.17	11.02	11.36
	0.10	0.74	1.46	2.42	3.28	4.40	5.36	6.17	7.27	8.00	9.24	10.00
	0.30	0.49	1.11	1.83	2.50	3.20	3.94	4.62	5.39	6.04	6.86	7.60
	0.50	0.37	0.94	1.52	2.06	2.68	3.32	7.88	4.52	5.09	5.81	6.35
51	0.05	0.74	1.76	2.82	3.95	4.78	6.29	7.53	8.50	9.70	10.77	11.07
	0.10	0.60	1.43	2.44	3.45	4.24	5.37	6.42	7.57	8.33	9.57	10.23
	0.30	0.41	1.09	1.81	2.59	3.25	4.25	4.86	5.69	6.31	6.99	7.81
	0.50	0.33	0.91	1.56	2.21	2.79	3.46	4.04	4.81	5.27	5.89	6.62

Table 9. The Critical Values for Two Point Detection about the Data Sets A and C

data set	$\alpha \backslash w$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
A	0.05	2.98	3.15	3.56	4.19	5.25	6.46	6.90	7.92	9.16	9.76	11.55
	0.10	2.29	2.53	2.95	3.56	4.50	5.18	6.05	6.83	7.38	8.25	9.89
	0.30	1.27	1.68	2.08	2.59	3.22	3.77	4.30	5.10	5.48	6.09	7.07
	0.50	0.92	1.33	1.70	2.15	2.61	3.10	3.57	4.10	4.50	5.03	5.82
C	0.05	1.99	2.56	3.27	4.27	5.29	6.04	6.89	7.88	8.88	9.81	11.58
	0.10	1.64	2.17	2.81	3.55	4.49	5.23	6.05	7.00	7.45	8.42	9.99
	0.30	1.13	1.57	2.10	2.62	3.23	3.75	4.39	5.05	5.49	6.12	6.99
	0.50	0.91	1.30	1.73	2.15	2.66	3.06	3.56	4.19	4.53	4.96	5.91

Table 10. Critical Values for Two Point Detection for Number of Data Points 11 to 51.

number of point	$a \backslash w$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
11	0.05	4.56	5.25	6.12	7.67	8.85	9.49	10.63	12.21	14.16	14.00	17.16
	0.10	3.68	4.07	4.76	5.74	6.37	7.51	8.24	9.02	10.39	11.08	12.69
	0.30	2.23	2.56	2.98	3.55	3.88	4.52	5.11	5.63	6.01	6.88	7.76
	0.50	1.68	1.98	2.23	2.70	3.03	3.57	3.79	4.17	4.61	5.25	5.57
21	0.05	2.49	2.86	3.42	4.23	5.27	6.25	6.90	7.9	9.02	9.79	11.57
	0.10	1.97	2.35	2.88	3.56	4.50	5.21	6.05	6.92	7.42	8.34	9.94
	0.30	1.2	1.63	2.10	2.61	3.23	3.76	4.35	5.08	5.49	6.11	7.03
	0.50	0.92	1.32	1.72	2.15	2.64	3.08	3.57	4.15	4.52	5.00	5.87
31	0.05	1.38	1.91	2.61	3.71	4.39	5.81	6.75	7.44	9.18	9.63	10.42
	0.10	1.17	1.65	2.35	3.31	3.97	5.11	5.72	6.58	7.67	8.12	9.21
	0.30	0.85	1.30	1.89	2.52	3.10	3.76	4.42	5.12	5.63	6.40	7.00
	0.50	0.70	1.13	1.61	2.09	2.60	3.16	3.78	4.36	4.80	5.27	5.94
41	0.05	1.16	1.75	2.56	3.48	4.67	5.61	6.47	7.50	8.71	10.04	10.46
	0.10	0.99	1.53	2.27	3.14	4.01	4.97	5.78	6.63	7.61	8.65	9.21
	0.30	0.70	1.20	1.82	2.40	3.15	3.83	4.46	5.15	5.89	6.61	7.22
	0.50	0.57	1.04	1.55	2.10	2.70	3.26	3.84	4.41	5.07	5.67	6.25
51	0.05	0.90	1.65	2.53	3.59	4.32	5.64	6.60	7.69	8.30	9.60	10.84
	0.10	0.81	1.43	2.23	3.14	3.88	5.04	5.93	6.85	7.59	8.52	9.52
	0.30	0.61	1.14	1.76	2.53	3.15	3.95	4.70	5.46	6.00	6.78	7.44
	0.50	0.5	0.99	1.56	2.18	2.74	3.44	3.99	4.62	5.16	5.86	6.48

REFERENCES

1. ANDREWS, D.F., and PREGIBON, D. (1978), "*Finding the Outliers That Matter*," Journal of the Royal Statistical Society, Ser. B, 40, 87-93.
2. BELSLEY, D.A., KUH, E., and WELSCH, R.E. (1980), Regression Diagnostics, New York ; John Wiley.
3. COCK, R.D. (1977), "*Detection of Influential Observations in Linear Regression*," Technometrics, 19, 15-18.
4. COOK, R.D., and WEISBERG, S. (1980), "*Characterizations of an Empirical Influence Function for Detecting Influential Cases in Regression*," Technometrics, 22, 495-508.
5. GENTLEMAN, J.F. and WILK, M.B. (1975 a), "*Detecting Outliers in a Two-Way Table : I. Statistical Behavior of Residuals*," Technometrics, 17, 1-14.
6. _____ (1975 b), "*Detecting Outliers* II. *Supplementing the Direct Analysis of Residuals*," Biometrics, 31, 387-410.
7. MICKEY, M.R., DUNN, O.J., and CLARK, V. (1976), "*Note on the Use of Stepwise Regression in Detecting Outliers*," Computers and Biomedical Research, 1, 105-111.