Communications for Statistical Applications and Methods
2017, Vol. 24, No. 3, 317–324

https://doi.org/10.5351/CSAM.2017.24.3.317
Print ISSN 2287-7843 / Online ISSN 2383-4757

A cautionary note on the use of Cook's distance

Myung Geun Kim[1,a]

[a]Department of Mathematics Education, Seowon University, Korea

Abstract

An influence measure known as Cook's distance has been used for judging the influence of each observation on the least squares estimate of the parameter vector. The distance does not reflect the distributional property of the change in the least squares estimator of the regression coefficients due to case deletions: the distribution has a covariance matrix of rank one and thus it has a support set determined by a line in the multidimensional Euclidean space. As a result, the use of Cook's distance may fail to correctly provide information about influential observations, and we study some reasons for the failure. Three illustrative examples will be provided, in which the use of Cook's distance fails to give the right information about influential observations or it provides the right information about the most influential observation. We will seek some reasons for the wrong or right provision of information.

Keywords: case deletion, Cook's distance, influence, regression

1. Introduction

In a regression context, there are many measures of the influence of observations on the least squares estimate of the parameter vector. Some of them can be found in Chatterjee and Hadi (1988) and Cook and Weisberg (1982). Since the change in the least squares estimate of regression coefficients due to case deletions is a vector quantity, it is usually normalized or scaled so that observations can be ordered in a certain way. In this vein Cook (1977) introduced an influence measure based on confidence ellipsoids.

The distribution of the change in the least squares estimator of the regression coefficients due to case deletions has a support set determined by a line in the multidimensional Euclidean space. Cook's distance does not reflect this kind of distributional property, and thus it can reduce or enlarge the influence of an observation on the least squares estimate. As a result, the use of Cook's distance is likely to underestimate the influence of an observation on the least squares estimate or overestimate it, which will be studied in Section 2. Hence the use of Cook's distance may lead to a wrong detection of influential observations. In Section 3 we consider three illustrative examples. For two examples, the use of Cook's distance fails to give the right information about influential observations, and for one example, it provides the right information about the most influential observation. We will seek some reasons for the wrong or right provision of information.

[1] Department of Mathematics Education, Seowon University, 377-3, Musimseo-ro, Seowon-gu, Cheongju-si, Chungcheongbuk-do 28674, Korea. E-mail: mgkim@seowon.ac.kr

Published 31 May 2017 / journal homepage: http://csam.or.kr

2. On Cook's distance

A linear regression model of our interest can be expressed as

$$y = X\beta + \varepsilon,$$

where y is an $n \times 1$ vector of values of the response variable, $X = (x_1, \ldots, x_n)^T$ is an $n \times p$ matrix of full column rank consisting of n measurements on the p fixed explanatory variables possibly including the intercept term, $\beta = (\beta_0, \beta_1, \ldots, \beta_{p-1})^T$ is a $p \times 1$ vector of unknown regression coefficients, and ε is an $n \times 1$ vector of independent random errors, each of which has zero mean and unknown variance σ^2. We write the least squares estimator of β as $\hat{\beta} = (X^T X)^{-1} X^T y$ which is an unbiased estimator of β and whose covariance matrix is given by $\text{cov}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$. The $n \times n$ matrix $H = (h_{ij}) = X(X^T X)^{-1} X^T$ is the hat matrix, and $e = (e_1, \ldots, e_n)^T = (I - H)y$ is the residual vector. The mean of the residual vector e is zero and its covariance matrix is $\text{cov}(e) = \sigma^2 (I - H)$. An unbiased estimator of σ^2 is given by $\hat{\sigma}^2 = e^T e/(n - p)$. More details can be found in Seber (1977).

The least squares estimator of β computed without observation r is written as $\hat{\beta}_{(r)}$. Miller (1974) showed that

$$\hat{\beta} - \hat{\beta}_{(r)} = \left(X^T X\right)^{-1} x_r \frac{e_r}{1 - h_{rr}}, \qquad r = 1, \ldots, n.$$

The mean vector of $\hat{\beta} - \hat{\beta}_{(r)}$ is zero and its covariance matrix is

$$\text{cov}\left(\hat{\beta} - \hat{\beta}_{(r)}\right) = \frac{\sigma^2}{1 - h_{rr}} \left(X^T X\right)^{-1} x_r x_r^T \left(X^T X\right)^{-1}.$$

The rank of $\text{cov}(\hat{\beta} - \hat{\beta}_{(r)})$ is one for nonnull x_r. The only nonzero eigenvalue of $\text{cov}(\hat{\beta} - \hat{\beta}_{(r)})$ is $\sigma^2 x_r^T (X^T X)^{-2} x_r/(1 - h_{rr})$ and its associated eigenvector is $(X^T X)^{-1} x_r$. When we denote by V_r a one-dimensional subspace generated by $(X^T X)^{-1} x_r$ of the p-dimensional Euclidean space, the subspace V_r is just a line along which the eigenvector $(X^T X)^{-1} x_r$ lies, and each $\hat{\beta} - \hat{\beta}_{(r)}$ has a distribution with which a random variable takes on values in the set V_r with probability one. More details about the distribution of $\hat{\beta} - \hat{\beta}_{(r)}$ can be found in Kim (2015).

In order to investigate the change in the value of $\hat{\beta}$ due to a deletion of observation r, Cook (1977) introduced an influence measure based on confidence ellipsoids as follows:

$$D_r = \frac{1}{p\hat{\sigma}^2} \left(\hat{\beta} - \hat{\beta}_{(r)}\right)^T \left(X^T X\right) \left(\hat{\beta} - \hat{\beta}_{(r)}\right).$$

Let $X^T X = GLG^T$ be the spectral decomposition of $X^T X$, where $L = \text{diag}(l_1, \ldots, l_p)$ is a $p \times p$ diagonal matrix consisting of the eigenvalues of $X^T X$, $G = (g_1, \ldots, g_p)$ is a $p \times p$ orthogonal matrix, and g_i is the eigenvector of $X^T X$ associated with the eigenvalue l_i. Then $X^T X = \sum_{i=1}^p l_i g_i g_i^T$. Hence D_r can be expressed as

$$\begin{aligned} D_r &= \frac{1}{p\hat{\sigma}^2} \sum_{i=1}^p l_i \left[\left(\hat{\beta} - \hat{\beta}_{(r)}\right)^T g_i\right]^2 \\ &= \frac{\left\|\hat{\beta} - \hat{\beta}_{(r)}\right\|^2}{p\hat{\sigma}^2} \sum_{i=1}^p l_i \cos^2 \theta_{ri}, \qquad (2.1) \end{aligned}$$

Table 1: The eigenvalues of X^TX

l_1	l_2	l_3	l_4	l_5
44676.2060	5965.4221	809.9521	105.4187	0.0012

where θ_{ri} is the angle between $\hat{\beta} - \hat{\beta}_{(r)}$ and g_i, and $\|\hat{\beta} - \hat{\beta}_{(r)}\|$ is the Euclidean norm of $\hat{\beta} - \hat{\beta}_{(r)}$.

The set $\{g_1, \ldots, g_p\}$ forms an orthonormal basis for the p-dimensional Euclidean space. The coordinate of $\hat{\beta} - \hat{\beta}_{(r)}$ with respect to the i^{th} eigenvector g_i is $(\hat{\beta} - \hat{\beta}_{(r)})^T g_i$. In the light of equation (2.1) the terms l_i and $(\hat{\beta} - \hat{\beta}_{(r)})^T g_i$ for each r play a specific role in determining the magnitude of D_r. The adoption of X^TX for scaling the Euclidean norm of $\hat{\beta} - \hat{\beta}_{(r)}$ is not reasonable as explained in what follows. In real data analyses, the line V_r is not in general parallel to any of the eigenvectors $g_1, \ldots, g_p$. Also, even in the case where the line V_r is almost parallel to one of the g_i's, it is nearly orthogonal to the other eigenvectors: for example, if the line V_r is almost parallel to g_p, then the component of D_r associated with the eigenvector g_p, $l_p[(\hat{\beta} - \hat{\beta}_{(r)})^T g_p]^2/p\hat{\sigma}^2$ nearly makes a real contribution to the influence of observation r on $\hat{\beta}$, while the component of D_r associated with the remaining eigenvectors, $\sum_{i=1}^{p-1} l_i[(\hat{\beta} - \hat{\beta}_{(r)})^T g_i]^2/p\hat{\sigma}^2$ is likely to distort the influence of observation r on $\hat{\beta}$. Most or all of the terms $(\hat{\beta} - \hat{\beta}_{(r)})^T g_i$ for each observation are computed in the outside of the set V_r over which $\hat{\beta} - \hat{\beta}_{(r)}$ is distributed. As a result, the terms $(\hat{\beta} - \hat{\beta}_{(r)})^T g_i$ play a role of having the value of D_r reduced or enlarged depending on the values of l_i as compared with the real influence of observation r, which results in distorting the influence of observation r. Hence the use of D_r is likely to underestimate the influence of observation r on $\hat{\beta}$ or overestimate it, and the information about influential observations that the Cook's distance provides may not be reliable.

3. Three illustrative examples

We will apply the expressions of D_r in equation (2.1) to three data sets: the Hald data set (Draper and Smith, 1981) and the rat data set (Cook and Weisberg, 1982) which were analyzed also by Cook (1977), and the body fat data set (Neter *et al.*, 1996, p. 261). Using the probabilistic behavior of $\hat{\beta} - \hat{\beta}_{(r)}$ through the spectral decomposition of its covariance matrix $\mathrm{cov}(\hat{\beta} - \hat{\beta}_{(r)})$, Kim (2015) introduced an influence measure $M_r = x_r^T(X^TX)^{-2}x_r/(1 - h_{rr})$ to investigate the influence of deleting an observation on the least squares estimate $\hat{\beta}$, and the problem of deleting multiple cases was considered by Kim (2016). For these three data sets, the result based on the D_r values will be compared with that based on the M_r values.

3.1. Hald data

The regression model with the intercept term β_0 is fitted to the Hald data set which consists of 13 observations on a single dependent variable and four independent variables. The estimated regression coefficients are $\hat{\beta}_0 = 62.41, \hat{\beta}_1 = 1.55, \hat{\beta}_2 = 0.51, \hat{\beta}_3 = 0.10$, and $\hat{\beta}_4 = -0.14$.

For the values of D_r, observation 8 has the largest value $D_8 = 0.394$ and observation 3 has the second largest value $D_3 = 0.301$. Based on the D_r values, observation 8 is thus identified as the most influential observation. However, the influence measure M_r shows that observation 3 has the largest influence on the least squares estimate of β ($M_3 = 879.74$) but observation 8 is not significantly influential ($M_8 = 37.16$). In order to seek some reasons for which the two results are contradictory to each other, we will investigate sources of the D_r values for observations 3 and 8. The eigenvalues of X^TX and their associated eigenvectors are included in Tables 1 and 2, respectively.

(a) Each row in Table 3 shows a normalized vector of each $\hat{\beta} - \hat{\beta}_{(r)}$. The values of $\cos\theta_{ri}$ shown in

Table 2: The eigenvectors of $X^T X$

g_1	g_2	g_3	g_4	g_5
−0.01699	0.00372	0.00004	−0.01104	0.99979
−0.12789	−0.04278	−0.64590	−0.75134	−0.01028
−0.83968	−0.50922	−0.01812	0.18763	−0.01030
−0.19842	0.07211	0.75572	−0.61985	−0.01052
−0.48881	0.85653	−0.10665	0.12626	−0.01010

Table 3: Normalized vectors of $\hat{\beta} - \hat{\beta}_{(r)}$

Number					
1	−0.99982	0.00772	0.01084	0.00706	0.01158
2	0.99976	−0.01306	−0.01003	−0.01155	−0.00904
3	−0.99979	0.01002	0.01021	0.01065	0.01013
4	−0.99981	0.00799	0.01083	0.01008	0.00964
5	−0.99984	0.00158	0.01304	0.00142	0.01216
6	0.99982	−0.00691	−0.01005	−0.01004	−0.01052
7	0.99981	−0.00095	−0.01376	−0.01099	−0.00820
8	0.99969	−0.01306	−0.00889	−0.01656	−0.00986
9	0.99979	−0.01122	−0.01009	−0.00975	−0.01024
10	−0.99952	0.02344	0.00807	0.01604	0.00885
11	−0.99972	0.01230	0.00941	0.01523	0.00958
12	−0.99979	0.01058	0.01074	0.01015	0.00981
13	0.99981	−0.00893	−0.01118	−0.00874	−0.01007

Table 4: The values of $\cos\theta_{ri}$

r	i				
	1	2	3	4	5
1	−0.00016	0.00085	−0.00112	0.00436	−0.99999
2	−0.00019	0.00081	0.00090	0.00291	1.00000
3	0.00007	0.00010	0.00027	0.00010	−1.00000
4	0.00017	−0.00060	0.00119	0.00204	−1.00000
5	−0.00038	0.00009	−0.00152	0.01296	−0.99991
6	−0.00053	−0.00059	−0.00178	−0.00284	0.99999
7	0.00087	0.00295	−0.00653	−0.00713	0.99995
8	0.00025	−0.00084	−0.00282	0.00612	0.99998
9	−0.00014	−0.00013	0.00120	0.00024	1.00000
10	−0.00030	−0.00009	−0.00415	−0.01389	−0.99989
11	−0.00019	0.00027	0.00234	−0.00467	−0.99999
12	−0.00019	−0.00051	−0.00045	0.00006	−1.00000
13	0.00020	0.00054	0.00048	−0.00228	1.00000

Table 5: The ratios $\cos^2\theta_{8i}/\cos^2\theta_{3i}$

	i		
1	2	3	4
12.8	73.6	109.5	3,955

Table 4 can be considered as a measure of closeness of $\hat{\beta} - \hat{\beta}_{(r)}$ to the i^{th} eigenvector g_i of $X^T X$. As $\hat{\beta} - \hat{\beta}_{(r)}$ gets close to g_i, the value of $\cos\theta_{ri}$ approaches to one. We note from Tables 2 and 3 that both vectors $\hat{\beta} - \hat{\beta}_{(3)}$ and $\hat{\beta} - \hat{\beta}_{(8)}$ are almost parallel to the last eigenvector g_5 of $X^T X$, which can be confirmed by the $\cos\theta_{ri}$ values shown in the last column of Table 4. The second to fifth columns of Table 4 show that both vectors $\hat{\beta} - \hat{\beta}_{(3)}$ and $\hat{\beta} - \hat{\beta}_{(8)}$ are almost orthogonal to each of the eigenvectors $g_1, \ldots, g_4$ of $X^T X$.

(b) For observations 3 and 8, the ratios $\cos^2\theta_{8i}/\cos^2\theta_{3i}$ listed in Table 5 are much larger than one

Table 6: Two components of Cook's distance for observations 3 and 8

r	D_r	$\sum_{i=1}^{4} l_i \left[\left(\hat{\beta} - \hat{\beta}_{(r)}\right)^T g_i\right]^2 / p\hat{\sigma}^2$	$l_5 \left[\left(\hat{\beta} - \hat{\beta}_{(r)}\right)^T g_5\right]^2 / p\hat{\sigma}^2$
3	0.301	0.065	0.236
8	0.394	0.368	0.026

Table 7: The change in $\hat{\beta}_k$ due to deletion of each of observations 3 and 8

r	$\hat{\beta}_0 - \hat{\beta}_{0(r)}$	$\hat{\beta}_1 - \hat{\beta}_{1(r)}$	$\hat{\beta}_2 - \hat{\beta}_{2(r)}$	$\hat{\beta}_3 - \hat{\beta}_{3(r)}$	$\hat{\beta}_4 - \hat{\beta}_{4(r)}$
3	−76.18	0.76	0.78	0.81	0.77
8	25.16	−0.33	−0.22	−0.42	−0.25

for all $i = 1, 2, 3, 4$, and hence they show that observation 8 is located closer to all of four axes $g_1, \ldots, g_4$ than observation 3.

(c) Since the line V_3 over which $\hat{\beta} - \hat{\beta}_{(3)}$ is distributed is almost parallel to the eigenvector g_5 of $X^T X$, the component of D_3 associated with the eigenvector g_5 which is $l_5[(\hat{\beta} - \hat{\beta}_{(3)})^T g_5]^2 / p\hat{\sigma}^2$ nearly makes a real contribution to the influence of observation 3 on $\hat{\beta}$. These components of D_r ($r = 3, 8$) are included in Table 6. Also, the proportion of $l_5[(\hat{\beta} - \hat{\beta}_{(r)})^T g_5]^2 / p\hat{\sigma}^2$ to D_r is about 79% for observation 3 and about 7% for observation 8. On the other hand, the line V_3 is almost orthogonal to all of the eigenvectors $g_1, \ldots, g_4$ of $X^T X$, and therefore the component of D_3 associated with the eigenvectors $g_1, \ldots, g_4$ which is $\sum_{i=1}^{4} l_i[(\hat{\beta} - \hat{\beta}_{(3)})^T g_i]^2 / p\hat{\sigma}^2$ is likely to distort the influence of observation 3 on $\hat{\beta}$. Observation 8 can be interpreted similarly to observation 3. The difference of $\sum_{i=1}^{4} l_i[(\hat{\beta} - \hat{\beta}_{(r)})^T g_i]^2 / p\hat{\sigma}^2$ between observations 3 and 8 is approximately −0.303, while the difference of $l_5[(\hat{\beta} - \hat{\beta}_{(r)})^T g_5]^2 / p\hat{\sigma}^2$ between observations 3 and 8 is approximately 0.210. The extent that the distance D_8 distorts the influence of observation 8 on $\hat{\beta}$ is far more severe than that of D_3. Thus the component $\sum_{i=1}^{4} l_i[(\hat{\beta} - \hat{\beta}_{(r)})^T g_i]^2 / p\hat{\sigma}^2$ plays a role of making the value of D_8 large, while it plays a role of making the value of D_3 relatively small. Hence the distance D_8 enlarges the influence of observation 8 on $\hat{\beta}$, while the distance D_3 reduces the influence of observation 3.

This is a reason for which the use of the D_r values identifies observation 8 that is not significantly influential as the most influential one and it cannot detect observation 3 as the most influential one.

Even though the D_r value was introduced as an overall measure of the combined influence of observation r on all of the estimated regression coefficients, it would be desirable if the use of the D_r values reveals influential observations for each regression coefficient, but the use of the D_r values does not. The use of the D_r values asserts that deletion of observation 8 has the largest change in $\hat{\beta}$. However, deletion of observation 8 does not bring about a significant change in either estimated regression coefficient, while deletion of observation 3 has the largest change in all of the estimated regression coefficients, as can be seen in what follows. Numerical computations of the values $\hat{\beta}_k - \hat{\beta}_{k(r)}$, $k = 0, 1, \ldots, 4$; $r = 1, \ldots, 13$ show that deletion of observation 3 has the largest change in $\hat{\beta}_k$ for all $k = 0, 1, \ldots, 4$. Table 7 shows the change in $\hat{\beta}_k$ due to deletion of each of observations 3 and 8. After removal of observation 8 from the sample, numerical computations based on the remaining sample of size 12 show that deletion of observation 3 still has the largest change in $\hat{\beta}_k$ for all $k = 0, 1, \ldots, 4$ as listed in Table 8. After removal of observation 3 from the sample, numerical computations based on the remaining sample of size 12 show that deletion of observation 4 has the largest change −75.77 in $\hat{\beta}_0$, deletion of observation 11 has the largest change 0.77 in $\hat{\beta}_1$, deletion of observation 4 has the largest change 0.79 in $\hat{\beta}_2$, deletion of observation 11 has the largest change 0.90 in $\hat{\beta}_3$, and deletion of

Table 8: The change in $\hat{\beta}_k$ due to deletion of observations 3 after removal of observation 8 from the sample

$\hat{\beta}_0 - \hat{\beta}_{0(r)}$	$\hat{\beta}_1 - \hat{\beta}_{1(r)}$	$\hat{\beta}_2 - \hat{\beta}_{2(r)}$	$\hat{\beta}_3 - \hat{\beta}_{3(r)}$	$\hat{\beta}_4 - \hat{\beta}_{4(r)}$
−49.17	0.50	0.50	0.54	0.50

observation 4 has the largest change 0.75 in $\hat{\beta}_4$. We note that the M_r values provide useful information about influential observations for each regression coefficient.

3.2. Body fat data

We fit the regression model with the intercept term β_0 to the the body fat data set which has 20 measurements on a single dependent variable and three independent variables. The least squares estimates of the regression coefficients are $\hat{\beta}_0 = 117.08, \hat{\beta}_1 = 4.33, \hat{\beta}_2 = -2.86$, and $\hat{\beta}_3 = -2.19$.

Observation 3 has the largest value $D_3 = 0.299$ and observation 1 has the second largest distance $D_1 = 0.279$. The D_r values assert that observation 3 is the most influential observation. However, for the M_r values, observation 1 has the largest value $M_1 = 401.19$ and observation 3 has $M_3 = 150.22$, not the second largest value. We have contradictory results also for the body fat data. We will seek some reasons for this contradictory results by investigating sources of the D_r values for observations 1 and 3. Detailed computations will not be included here. The four eigenvalues of $X^T X$ are 81290.24, 294.25, 119.82, 0.00062. The eigenvector corresponding to the last eigenvalue is $(0.99909, 0.03012, -0.02583, -0.01592)$. Euclidean norm $\|\hat{\beta} - \hat{\beta}_{(r)}\|$ is 72.92 for observation 1 and 37.47 for observation 3.

(a) An investigation of the closeness between a normalized vector of each $\hat{\beta} - \hat{\beta}_{(r)}$ and each eigenvector of $X^T X$ shows that $\cos\theta_{r4}$ is -0.9999988 for observation 1 and 0.9999891 for observation 3, which implies that both vectors $\hat{\beta} - \hat{\beta}_{(1)}$ and $\hat{\beta} - \hat{\beta}_{(3)}$ are almost parallel to the last eigenvector g_4 of $X^T X$. Also, both vectors $\hat{\beta} - \hat{\beta}_{(1)}$ and $\hat{\beta} - \hat{\beta}_{(3)}$ are almost orthogonal to each of the remaining eigenvectors of $X^T X$.

(b) For observations 1 and 3, the ratio $\cos^2\theta_{3i}/\cos^2\theta_{1i}$ is 5.09, 5.18, 15.26 for $i = 1, 2, 3$, respectively. Hence observation 3 is located closer to all of three axes g_1, g_2, g_3 than observation 1.

(c) In the light of the results in (a), among the components of D_r $(r = 1, 3)$ given in the first expression of equation (2.1), only the component

$$\frac{l_4\left[\left(\hat{\beta} - \hat{\beta}_{(r)}\right)^T g_4\right]^2}{p\hat{\sigma}^2}$$

nearly makes a real contribution to the influence of observation r on $\hat{\beta}$, and its value is 0.133 for observation 1 and 0.035 for observation 3. Also, the proportion of $l_4[(\hat{\beta} - \hat{\beta}_{(r)})^T g_4]^2/p\hat{\sigma}^2$ to D_r is about 48% for observation 1 and about 12% for observation 3. On the other hand, since the line V_r $(r = 1, 3)$ is almost orthogonal to all of the remaining eigenvectors g_1, g_2, g_3 of $X^T X$, the component of D_r associated with the eigenvectors g_1, g_2, g_3 which is

$$\frac{1}{p\hat{\sigma}^2}\sum_{i=1}^{3} l_i\left[\left(\hat{\beta} - \hat{\beta}_{(r)}\right)^T g_i\right]^2$$

> is likely to distort the influence of observation r on $\hat{\beta}$. The component $\sum_{i=1}^{3} l_i[(\hat{\beta}-\hat{\beta}_{(r)})^T g_i]^2$ is 0.146 for observation 1 and 0.264 for observation 3. The extent that the distance D_3 distorts the influence of observation 3 on $\hat{\beta}$ is more severe than that of D_1. The component $\sum_{i=1}^{3} l_i[(\hat{\beta}-\hat{\beta}_{(r)})^T g_i]^2/p\hat{\sigma}^2$ plays a role of making the value of D_3 large, while it plays a role of making the value of D_1 relatively small. Hence the distance D_3 enlarges the influence of observation 3 on $\hat{\beta}$, while the distance D_1 reduces the influence of observation 1.

This is a reason why observation 3 has the largest D_r value, $D_3 = 0.299$, though it is not identified as a significantly influential observation by the M_r values.

Furthermore, the D_r values do not provide useful information about influential observations for each regression coefficient but the M_r values do as can be seen in what follows. Numerical computations of the values $\hat{\beta}_k - \hat{\beta}_{k(r)}$, $k = 0, 1, 2, 3$; $r = 1, \ldots, 20$ show that deletion of observation 1 has the largest change in $\hat{\beta}_k$ for all $k = 0, 1, 2, 3$: $\hat{\beta}_0 - \hat{\beta}_{0(r)}$ is -72.86 for observation 1 and 37.44 for observation 3, $\hat{\beta}_1 - \hat{\beta}_{1(r)}$ is -2.12 for observation 1 and 1.02 for observation 3, $\hat{\beta}_2 - \hat{\beta}_{2(r)}$ is 1.88 for observation 1 and -0.87 for observation 3, $\hat{\beta}_3 - \hat{\beta}_{3(r)}$ is 1.08 for observation 1 and -0.69 for observation 3. Observation 3 is identified as the most influential one by the D_r values but it does not have a significant influence on any estimate $\hat{\beta}_k$ ($k = 0, 1, 2, 3$).

3.3. Rat data

The regression model with the intercept term β_0 is fitted to the rat data set which consists of 19 measurements on a single dependent variable and three independent variables. The least squares estimates of the regression coefficients are $\hat{\beta}_0 = 0.27$, $\hat{\beta}_1 = -0.02$, $\hat{\beta}_2 = 0.01$, and $\hat{\beta}_3 = 4.18$.

For the D_r values, observation 3 has the largest value $D_3 = 0.930$. For the M_r values, observation 3 has the largest value $M_3 = 1864.3$. Both influence measures lead to the same conclusion that observation 3 is the most influential one. We will briefly seek some reasons for the same conclusion. Detailed computations will not be included here. The four eigenvalues of X^TX are 565097.6, 20.5, 0.16, 0.003. The eigenvector g_4 corresponding to the last eigenvalue is $(0.0213, -0.0052, 0.0005, 0.9998)$. For observation 3, we have Euclidean norm $\|\hat{\beta} - \hat{\beta}_{(3)}\| = 2.684$.

The cosine of the angle between $\hat{\beta} - \hat{\beta}_{(3)}$ and g_4 is 0.9993, which implies that $\hat{\beta} - \hat{\beta}_{(3)}$ is almost parallel to the last eigenvector g_4 of X^TX and that it is almost orthogonal to each of the remaining eigenvectors of X^TX. Hence, among the components of D_3, only the component $l_4[(\hat{\beta}-\hat{\beta}_{(3)})^T g_4]^2/p\hat{\sigma}^2$ nearly makes a real contribution to the influence of observation 3 on $\hat{\beta}$, and its value is 0.781. The component $\sum_{i=1}^{3} l_i[(\hat{\beta} - \hat{\beta}_{(3)})^T g_i]^2$ is 0.149. Also, the proportion of $l_4[(\hat{\beta} - \hat{\beta}_{(3)})^T g_4]^2/p\hat{\sigma}^2$ to D_3 is about 84% and it is very high. Therefore the extent that the distance D_3 reflects the real influence of observation 3 on $\hat{\beta}$ is very high so that the value D_3 can yield the same result as the value M_3.

4. Concluding remarks

The distance D_r is defined by scaling $\hat{\beta} - \hat{\beta}_{(r)}$ using the matrix X^TX. Almost all of the eigenvectors X^TX are not in general parallel to the line V_r over which $\hat{\beta} - \hat{\beta}_{(r)}$ is distributed. The distance D_r inevitably includes the component associated with the axes other than the axis determined by the line V_r. This component of D_r can be a source of distorting the influence of observation r on $\hat{\beta}$. Hence the information about influential observations that the Cook's distance provides may not be reliable. The first two examples analyzed in the previous section show defects of the distance D_r as an influence measure, while the three examples show that the M_r values can be a useful influence measure.

References

Chatterjee S and Hadi AS (1988). *Sensitivity Analysis in Linear Regression*, Wiley, New York.

Cook RD (1977). Detection of influential observation in linear regression, *Technometrics*, **19**, 15–18.

Cook RD and Weisberg S (1982). *Residuals and Influence in Regression*, Chapman and Hall, New York.

Draper NR and Smith H (1981). *Applied Regression Analysis* (2nd ed), Wiley, New York.

Kim MG (2015). Influence measure based on probabilistic behavior of regression estimators, *Computational Statistics*, **30**, 97–105.

Kim MG (2016). Deletion diagnostics in fitting a given regression model to a new observation, *Communications for Statistical Applications and Methods*, **23**, 231–239.

Miller RG (1974). An unbalanced jackknife, *Annals of Statistics*, **2**, 880–891.

Neter J, Kutner MH, Nachtsheim CJ, and Wasserman W (1996). *Applied Linear Regression Models* (3rd ed), McGraw-Hill, Irwin.

Seber GAF (1977). *Linear Regression Analysis*, Wiley, New York.

Received April 6, 2017; Revised May 6, 2017; Accepted May 8, 2017