THE DETECTION OF INFLUENTIAL SUBSETS IN LINEAR REGRESSION USING AN INFLUENCE MATRIX

BY
DANIEL PEÑA AND VICTOR J. YOHAI

SUMMARY

This paper presents a new method to identify influential subsets in linear regression problems. The procedure uses the eigenvalues of an influence matrix which is defined as the uncentered covariance of a set of vectors which represent the changes on the fit produced by the deletion of each point. This matrix is normalized to include the univariate Cook's statistics in the diagonal. It is shown that points in an influential subset will appear with large weights in at least one of the eigenvectors linked to the largest eigenvalues in this influence matrix. The method is illustrated with several well-known examples in the literature, and in all of them it succeeds in identifying the relevant influential subsets.

Key words: Diagnostics; Influential Observations; Masking; Multiple Outliers.

Authors footnote. Daniel Peña is Professor of Statistics, Universidad Carlos III de Madrid, 28903 Getafe, Madrid, Spain and Víctor J. Yohai is Professor of Statistics, Universidad de Buenos Aires, and CEMA. He is also a Resarcher at CONICET, Buenos Aires, Argentina. This research was supported by the Dirección General de Política Científica, MEC, Grant PB 87-0808.

1. INTRODUCTION

Many procedures are available to identify a single outlier or an isolated influential point in linear regression. Beckman and Cook (1983) and Chatterjee and Hadi (1986) survey some of these procedures. The detection of influential subsets or multiple outliers is more difficult, because the masking and swamping problems. Masking occurs when one outlier is not detected because of the presence of others; swamping when a non–outlier is wrongly identified due to the effect of some hidden outliers.

The procedures for dealing with multiple outliers or influential subset could be classified in four groups. The first includes sequential methods which are designed to avoid the masking problem. Marasinghe (1985) and Kianifard and Swallow (1989, 1990) have suggested a sequential testing strategy to identify a set of k points, where the maximum number of outliers in the sample, k, must be fixed in advance. The main weakness of these procedures is to be very sensitive to the choice of k, because the exact number of outliers is almost never known. The second group of methods are based on extensive checking of a large number of subsets, and includes the procedure proposed by a Cook and Weisberg (1982), among others. Although these methods are attractive, the computational burden involved made them not suited to analyze samples of medium or large size. The third group of methods are based on robust estimation. For instance, Rousseeuw and Leroy (1987) and Rousseeuw and Zomeren (1990) have suggested to overcome the masking problem by using robust estimates with high breakdown for the regression parameters. These estimates are computed using a resampling scheme. Hawkins, Bradu and Kass (1984) have proposed a diagnostic procedure which is also based on a resampling scheme. These procedures have proved to be very effective in dealing with masking problems, however they require extensive computations which become prohivitive when the number of carriers is large. Finally, the fourth group of techniques try to identify influential subsets by looking at the multivariate structure of the data points. Gray and Ling (1984) proposed the use of cluster analysis over a modified hat matrix to identify influential sets, and Hocking (1984) has suggested to compute the eigenstructure of the matrices $X'X$ and $(X\mathbf{y})'(X\mathbf{y})$ where $\mathbf{y}$ is the vector of responses and the matrix X contains the explanatory variables.

In this paper we present a new method to identify influential subsets by looking at the eigenvalues of an "influence matrix". This matrix is defined as the uncentered covariance of a set of vectors which represent the effect on the fit of the deletion of each data point. This matrix is normalized to have the univariate Cook's statistics in the diagonal. The method seems to work very well in all the data sets in which it has been tested.

The paper is organized as follows. Section 2 defines the influence matrix. Section 3 gives an heuristic justification of why the eigenvectors linked to non–null eigenvalues can be used to identify influential subsets. Section 4 applies the procedure to several examples.

2.THE INFLUENCE MATRIX

Consider a linear regression model between an independent variable Y and p carriers $X_1, ..., X_p$, and suppose that there are n data points $(y_i, x_{i1}, ...x_{ip})$, $1 \leq i \leq n$.

The following notation will be used in the rest of the paper: $\mathbf{y} = (y_1, ..., y_n)'$, $\mathbf{x}_i = (x_{i1}, ..., x_{ip})'$, X is the $n \times p$ matrix with rows $\mathbf{x}_1', ..., \mathbf{x}_n'$, Then according to the standard

linear model assumptions,

$$\mathbf{y} = X\mathbf{b} + \epsilon,$$

where $\mathbf{b} = (b_1, ..., b_p)'$ is the vector of regression coefficients and the vector of regression errors $\epsilon = (\epsilon_1, ..., \epsilon_n)'$, where the ϵ_i's are independent random variables with distribution $N(0, \sigma^2)$.

The least squares estimate (LSE) of $\mathbf{b}$ is given by

$$\hat{\mathbf{b}} = (X'X)^{-1}X'\mathbf{y},$$

the vector of fitted values $\hat{\mathbf{y}} = (\hat{y}_1, ..., \hat{y}_n)'$ by

$$\hat{\mathbf{y}} = X\hat{\mathbf{b}} = H\mathbf{y},$$

where $H = X(X'X)^{-1}X'$ is the hat matrix , and the vector of residuals $\mathbf{e} = (e_1, ..., e_n)'$ by

$$e_i = \mathbf{y} - X\hat{\mathbf{b}} = (I - H)\mathbf{y}.$$

Let $\hat{\mathbf{b}}_{(i)}$ be the LSE when the i-th data point is deleted, then the change in the LSE is given by (see Cook and Weisberg,1982, page 110)

$$\hat{\mathbf{b}} - \hat{\mathbf{b}}_{(i)} = \frac{e_i(X'X)^{-1}\mathbf{x}_i}{1 - h_{ii}}, \tag{1}$$

where h_{ij} is the ij-th element of H. Consequently if we denote by $\hat{y}_{j(i)}$ the new fitted value for observation j, we get

$$\hat{y}_j - \hat{y}_{j(i)} = \frac{h_{ij}e_i}{1 - h_{ii}}. \tag{2}$$

Masking occurs when there are several influential data points which produce similar effect on the least squares fit. In this case, the deletion of just one of them does not produce much change on the fit, and this explains why the procedures based on single delation fail in detecting this type of influential sets.

Put $\hat{\mathbf{y}}_{(i)} = (\hat{y}_{1(i)}, ..., \hat{y}_{n(i)})'$, then the vector $\mathbf{t}_i = \hat{\mathbf{y}} - \hat{\mathbf{y}}_{(i)}$ summarizes the effect on the fit of deleting the observation i-th.

We will say that two observations i and j have similar effects on the least squares fit when $\mathbf{t}_i \approx \lambda\mathbf{t}_j$ for some scalar $\lambda > 0$ and opposed effects when $\lambda < 0$. Then, in order to detect possible sets of influential observations having similar or opposed effect on the fit, it seems plausible to look at the uncentered covariance matrix of the $\mathbf{t}_i$'s. Therefore, we define the $n \times n$ influence matrix M as the normalized version of this covariance matrix given by

$$M = [\text{illegible}] = \frac{1}{p\,s^2} \begin{pmatrix} t_1't_1 & \cdots & t_1't_n \\ \vdots & & \vdots \\ t_n't_1 & & t_n't_n \end{pmatrix}$$

where $s^2 = \sum_{i=1}^{n} \epsilon_i^2/(n-p)$.

Using (2), and the fact that H is idempotent it is immediate to show that if we denote by m_{ij} the ij-th element of M, then

$$m_{ij} = \frac{e_i e_j h_{ij}}{(1-h_{ii})(1-h_{jj})ps^2}.$$

Since H is a semi positive definite matrix of rank p, M has this property too, except when either some e_i or some h_{ii} vanishes. Observe that the diagonal elements of M are the Cook's statistics.

3. A PROCEDURE FOR DETECTING INFLUENTIAL SETS

Let I be an index set corresponding to a subset of data points. Cook and Weisberg (1980) proposed to measure the joint influence of the data points with index in I by

$$D_I = \frac{(\hat{\mathrm{b}} - \hat{\mathrm{b}}_{(I)})' X' X (\hat{\mathrm{b}} - \hat{\mathrm{b}}_{(I)})}{ps^2},$$

where $\hat{\mathrm{b}}_{(I)}$ is the LSE computed after deletion of the data points with index in I.

It may be shown that this statistics can be written as

$$D_I = \frac{e_I'(I-H_I)^{-1}H_I(I-H_I)^{-1}e_I}{ps^2},$$

where the components of e_I are the least squares residuals and H_I the submatrix of H corresponding to the set I.

Theoretical influence curves (see Hampel, 1974) corresponding to infinitesimal fractions of outliers are linear. Since the empirical influence curve is given by the $n(\hat{\mathrm{b}}_{(i)} - \hat{\mathrm{b}})$'s and it converges to the theoretical one, it seem plausible to use the following linear approximation when the size of I is small relative to n

$$(\hat{\mathrm{b}} - \hat{\mathrm{b}}_{(I)}) \approx \sum_{i \in I} (\hat{\mathrm{b}} - \hat{\mathrm{b}}_{(i)}). \tag{3}$$

Using (1) and (3) we get the following approximation

$$D_I \approx C_I = \sum_{i \in I} \sum_{j \in I} m_{ij}. \tag{4}$$

Therefore as long as the approximation given by (4) holds, one way of detecting influential sets is by searching large values of C_I. This may be done for example using integer programming algorithms, however this alternative is not further pursued here.

In this paper we propose a procedure to detect sets I with large C_I based on the eigenvalues and eigenvectors of M. The following limit case will give an heuristic justification of the proposed procedure.

Let r_{ij} be the uncentered correlation coefficient between $\mathbf{t}_i$ and $\mathbf{t}_j$, then

$$r_{ij} = \frac{m_{ij}}{m_{ii}^{1/2} m_{jj}^{1/2}}.$$

Suppose that there are k groups of influential observations $I_1, ..., I_k$, such that

(i) If $i, j \in I_h$, then $|r_{ij}| = 1$. This means that the effects on the least squares fit produced by the deletion of two points in the same set I_h have correlation 1 or -1.

(ii) If $i \in I_j$ and $l \in I_h$ with $j \neq h$, then $r_{il} = 0$. This means that the effects produced on the least squares fit by observations i and j belonging to different sets are uncorrelated.

(iii)If i does not belong to any I_h, then $r_{ij} = 0$ for all j. This means that data points outside these groups have no influence on the fit. Then, according to (i) we can split each set I_h in I_h^1 and I_h^2 such that:

(1) If $i, j \in I_h^q$, then $r_{ij} = 1$

(2) If $i \in I_h^1$ and $j \in I_h^2$, then $r_{ij} = -1$

Let $\mathbf{v}_1 = (v_{11}, ..., v_{1n})', ..., \mathbf{v}_k = (v_{k1}, ..., v_{kn})'$ be defined by

$$v_{hj} = \begin{cases} m_{jj}^{1/2} & \text{if } j \in I_h^1 \\ -m_{jj}^{1/2} & \text{if } j \in I_h^2 \\ 0 & \text{if } j \notin I_h. \end{cases}$$

Then it is easy to show that if (i)-(iii) hold, then

$$M = \sum_{i=1}^{k} \mathbf{v}_i \mathbf{v}_i',$$

and since the $\mathbf{v}_i$'s are orthogonal, this implies that the eigenvectors of M are $\mathbf{v}_1, ..., \mathbf{v}_k$, and that the corresponding eigenvalues $\lambda_1, ..., \lambda_k$ are given by

$$\lambda_h = \sum_{i \in I_h} m_{ii}.$$

It is clear that when the matrix M satisfies (i)-(iii), the only sets I with large C_I are I_h^q, $1 \leq h \leq k$, $q = 1, 2$, and these sets may be found by looking at the eigenvectors associated to non-null eigenvalues of M.

For real data sets, (i)-(iii) do not hold exactly. However the masking effect is typically produced by the presence in the sample of blocks of influential observations producing similar or opposed effects. These blocks are likely to produce a matrix M with a structure close to the one described on (i)-(iii). In fact, two influential observations i, j producing similar effects should have r_{ij} close to 1, and close to -1 when they have opposed effects. Influential observations with non correlated effects have $|r_{ij}|$ close to 0. The same will happen with non influential observations. In this case the eigenvectors will have approximately the structure described above, and the null components will be replaced by small values.

This suggests the following procedure to identify influential sets:

(a)Find the eigenvectors corresponding to the p non-null eigenvalues of the influence matrix M.

(b)Consider the eigenvectors corresponding to large eigenvalues, and define the sets I_j^1 and I_j^2 by those components with large positive and negative weights respectively.

In Section 4 we apply this procedure to several examples where the methods based on individual deletion fail due to masking effects. In all the cases our procedure succeeds in detecting the influential sets.

4. EXAMPLES

Example 1. This first example is designed to show the interpretation of the eigenvectors of the influential matrix in three simple masking schemes (see table 1 and figure 1). In the three cases we have eight good points generated by $y = 1 + x + u$ where u is a normal random variable with mean 0 and standard deviation 0.1 and two high leverage points. In case (a) we have the standard masking scheme in which both outliers produce the same effect and one is masked by the other, in (b) the two outliers produce opposite effects, in (c) we have swamping, that is, the 9-th point appears as outlier because of the effect of the 10-th point.

Table 2 presents the largest eigenvalue of the influence matrix and the corresponding eigenvector in three cases. In case (a) the largest eigenvalue is roughly three times the next one and gives the largest weight to the two outliers. Also the two outliers have positive weight, whereas all the good points have a small and negative one. Therefore, the analysis shows the presence of two different sets of points. In case (b) the two outliers are again clearly identified: they appear in the eigenvector corresponding to the largest eigenvalue with large values and opposite sign, whereas the rest of the points are given zero weight. Finally, in case (c) the outlier is given a large and positive weight, whereas all the good points have negative weight, with the greatest value at the good high leverage point. In summary, the components of the eigenvector corresponding to the largest eigenvalue show in all cases the relevant structure of the data set.

(figure 1 about here)

(table 1 and 2 about here)

Example 2. As second example we consider the data of international phone calls in Belgium used by Rousseeuw and Leroy (1987). (See figure 2). The largest eigenvalue of the influence matrix is 1.16, 16.5 times greater than the second one. Its eigenvector (see table 3) gives a very small weight to the first fourteen good observations, large and negative weight to the six outliers and large and positive to the last four good points. The second eigenvector gives a negative value to the first fourteen data points and a positive value to the rest.

(figure 2 about here)

(table 3 about here)

In summary, the eigenvectors show that there are 6 outliers which behave very differently from all the other points. It is interesting to point out that a measure of univariate influence as Cook's D does not show any evidence of influential sets due to the masking effect: the largest values of this univariate statistics are rather small (see table (3) and correspond to point 20th, ($D = .27$) which is an outlier, and point 24th, ($D = .22$), which is not.

Table 4 includes the values of the multivariate test statistic and the F–value for the standard outlier test based on the decrease in the residual sum of square when the subset is deleted (see Barnett and Lewis, 1978, p. 265). As it is well known, this F observed value must be compared with the distribution of the maximum F over all sets of the same size, and this distribution is unknown (see Beckman and Cook (1983)). However, the large value of the F for the set {15, ..., 20} suggests that this set contains outliers. Note also that this set is very influential. These values are only identified as outliers according to the F test when the five points are deleted, due to the masking effect. On the other hand, both sets $\{15,...,20\}$ and $\{21,...,24\}$ are very influential, although figure 2 shows that the first includes outliers and the second good high leverage points. The swamping effect appears (table 4) in the value of the F statistics for set $\{21,...,24\}$, due to the presence of the outlier set $\{15,...,20\}$. When this later set is removed, the set $\{21,...,24\}$ is still influential, and, given the large value of the F statistics, observation 21 could be considered an outlier, whereas the other three points seem to be correct.

(table 4 about here)

Example 3. Data of the Hertzsprung–Russell diagram of a star cluster, from Rousseeuw and Leroy (1987). The data are plotted in figure 3, where four giant stars which correspond to points $\{11, 20, 30, 34\}$ can be seen as outliers. Table 5 shows the components of the eigenvector corresponding to the largest eigenvalue. These components are also plotted in figure 4. It can be seen that points 11, 20, 30 and 34 have a common and large effect. It is also shown that points $\{7, 14, 17\}$ seem to have some effect, specially 7 and 14, but with opposite effect than the others (see figure 3). Table 6 shows the values of the multiple Cook's D statistic and the F value for different combinations of points in these sets. Because of the masking effect we need to delete the four points $\{11, 20, 30, 34\}$ in order to see its joint effect clearly. The set $\{7, 14, 17\}$ is neither influential nor outlying.

(figure 3 about here)

(tables 5 and 6 about here)

(figure 4 about here)

Example 4. We will use the well-known stack-loss data from Daniel and Wood (1980). After a detailed search they identified points (1, 3, 4, 24) as outliers. Cook (1979) using a sequential search found (1, 2, 4, 21) as influential or outliers points. Gray and Ling using their k-clustering algorithm ended up with (1, 2, 3, 4, 21). Finally Rousseeuw and Zomeren (1990) also identified these five points.

Table 7 gives the two largest eigenvalues and the corresponding eigenvectors of the influence matrix. The first eigenvector is clearly dominated by the 21-th observation which receives a weight 3.5 times the next largest one. The second eigenvector gives largest weight to $\{1, 2, 3, 4\}$. Table 8 summarizes the results of deleting different combinations of these points. The most influential set is $\{1, 2, 3, 4\}$, which can also be considered a set of outliers, whereas point $\{21\}$ could be outlier, although it is not very influential.

(tables 7 and 8 about here)

Example 5. We use here the artificial data generated by Hawkins, Bradu and Kass (1984). The model contains 75 data points in four dimensions (one response and 3 explanatory variables). The first 10 data points are high leverage outliers, and the next four points are good observations with high leverage. The rest of the observations are good points with low leverage.

The eigenvalues of M are $\lambda_1 = 2.36$, $\lambda_2 = 1.63$, $\lambda_3 = 0.11$ and $\lambda_4 = 0.04$. The coefficients of the eigenvectors corresponding to λ_1 and λ_2 are shown in table 9 and figure 5.

The first eigenvector gives high positive weight to observations in the set $\{11, 13, 14\}$, specially to observation 14. All these points are good high leverage points.

Two sets of large coefficients may be distinguished in the second eigenvalue: the set $\{1, ..., 10, 14\}$ with negative coefficients, and the set $\{11, 12, 13\}$ with positive coefficients. Thus, the first set includes all the outliers and one good leverage point, and the second set three of the good high leverage points.

It may be observed in table 9 that the only large values of the univariate Cook's D statistic corresponds to good leverage points, and therefore they do not detect any outlier point.

Table 10 summarizes the results of deleting different combinations of these sets. Both sets, $I_1 = \{1, ..., 10\}$ and $I_2 = \{11, 12, 13, 14\}$ have very large D_I. However, once the observations in I_1 are deleted the F value for testing the set I_2 is not significant. Instead, once the observations in I_2 are deleted, the set I_1 is very influential, and the large F value suggest that their points are outliers.

(tables 9 and 10 about here)

(figure 5 about here)

REFERENCES

Barnett, V., and Lewis, T. (1978), *Outliers in Statistical Data*, John Wiley and Sons.

Beckman, R.J., and Cook, R.D. (1983), "Outlier...s," *Technometrics*, 25, 119–163.

Chatterjee, S., and Hadi A.S. (1986), "Influential Observations, High Leverage Points, and Outliers in Lineal Regression," *Statistical Science*, 1, 3, 379–416.

Cook, R.D. (1979), "Influential Observations in Linear Regression," *Journal of American Statistical Association*, 74, 169–174.

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

Daniel, C., and Wood, F.S. (1980), *Fitting Equations To Data*, John Wiley and Sons.

Gray, J.B., and Ling, R.F. (1984), "K–Clustering as a Detection Tool for Influential Subsets in Regression," *Technometrics*, 26, 305–330.

Hampel, F.R. (1974), "The Influence Curve and its Role in Robust Estimation," *Journal of American Statistical Association*, 69, 383–393.

Hawkins, D.M., Bradu, D. and Kass, G.V. (1984), "Location of Several Outliers in Multiple Regression Data Using Elemental Sets," *Technometrics*, 26, 197–208.

Hocking, R.R. (1984), "Discussion of Gray and Ling paper," *Technometrics*, 26, 321–323.

Kianifard, F., and Swallow, W. (1990), "A Monte Carlo Comparison of five Procedures for Identifying Outliers in Lineal Regression," *Communication in Statistics (Theory and Methods)*, 19, 1913–1938.

Mararinghe, M.G. (1985), "A Multistage Procedure for Detecting Several Outliers in Linear Regression," *Technometrics*, 27, 395–399.

Rousseeuw, P.J. and Zomeren, B.C. (1990), "Unmasking Multivariate Outliers and Leverage Points," *Journal of American Statistical Association*, 85, 633–651.

Rousseeuw, P.J. and Leroy, A.M. (1987), *Robust Regression and Outlier Detection*, John Wiley and Sons.

case	1	2	3	4	5	6	7	8	9(a)	10(a)	9(b)	10(b)	9(c)	10(c)
x	1	2	3	4	5	6	7	8	12	12	12	12	12	12
y	2.0	2.9	3.9	5.1	6.2	6.9	7.8	9.1	19	20	19	7	13	7

Table 1. Data for Example 1.

	λ_1	λ_1/λ_2	1	2	3	4	5	6	7	8	9	10
(a)	1.27	2.87	$-.17$	$-.06$	$-.00$	$-.00$	$-.02$	$-.10$	$-.22$	$-.33$	.42	.79
(b)	3.78	3.783	.00	$-.00$	$-.00$	$-.00$	$-.00$	.00	$-.00$	$-.00$	$-.71$	.71
(c)	3.25	32	$-.05$	$-.02$	$-.00$	$-.00$	$-.01$	$-.02$	$-.04$	$-.10$	$-.50$	.85

Table 2. Largest Eigenvalue, Ratio to the Next one and Eigenvector for Example 1.

	1-14	15	16	17	18	19	20	21	22	23	24
eigenvector	$-.002 < v < .07$	$-.13$	$-.15$	$-.20$	$-.26$	$-.35$	$-.48$	.21	.34	.38	.43
Cook D	$0 < d < .01$	.02	.03	.05	.08	.14	.27	.05	.13	.17	.22

Table 3. Elements of First Eigenvector and Cook's Statistics for the International Phone Call Data.

set	D	F
$\{15\}$	0.03	1.06
$\{15, 16\}$	0.12	1.13
$\{15, 16, 17\}$	0.40	1.48
$\{15, 16, 17, 18\}$	1.03	2.20
$\{15, 16, 17, 18, 19\}$	2.61	4.39
$\{15, 16, 17, 18, 19, 20\}$	6.74	5.48
$\{21, 22, 23, 24\}$	6.93	6.50
$\{21\}$	0.05	0.87
$\{22, 23, 24\}$	3.32	4.74
$\{21 \vert \{15, ..., 20\}\}$	1.43	130.14
$\{22, 23, 24 \vert \{15, ..., 20\}\}$	5.39	2.72
$\{21, 22, 23, 24 \vert \{15, ..., 20\}\}$	0.80	31.44

Table 4. Values of Cook's D for Multiple Cases and F Value for the International Phone Calls Data. The Notation $\{A|B\}$ Means that Set B is Completly Deleted from the Analysis of the Influence of Set A.

case	7	11	14	17	20	30	34
$\lambda_1 = 1,05$	.20	−.25	.28	.13	−.36	−.47	−.61
D	.04	.06	.09	.05	.14	.23	.41

Table 5. Eigenvector Coefficients Greater that .10 and Values of the D Statistic for the Hertzsprung-Russell Data of a Star Cluster.

set	D	F
$\{11, 20\}$	0.68	1.11
$\{30, 34\}$	2.22	3.95
$\{11, 20, 30, 34\}$	41.44	11.53
$\{7, 14\}$	.29	2.80
$\{7, 14, 17\}$	.52	3.75
$\{11, 20, 30, 34; 13, 14, 17\}$	33.71	7.21

Table 6. Cook's D Statistic for Multiple Cases and F Value for Outliers for the Hertzsprung-Russell Data.

Case	Eigenvalue Coefficients		D
	$\lambda_1 = .88$	$\lambda_2 = .39$	
1	−.21	−.51	.15
2	.12	.31	-
3	−.25	−.42	.13
4	.10	−.50	.13
5	-	-	-
6	-	.18	-
7	-	.20	-
8	-	.12	-
9	-	.22	-
10	-	-	-
11	−.15	.10	-
12	−.21	.18	-
13	-	-	-
14	-	-	-
15	-	-	-
16	-	-	-
17	-	-	-
18	-	-	-
19	-	-	-
20	-	-	-
21	.88	−.12	.69

Table 7. Two Largest Eigenvalues and its Eigenvectors and Univariate Cook's D for the Stack-Loss Data. Values with Absolute Value Smaller than .1 are Omitted.

set	D	F
$\{21\}$	.69	11.09
$\{1,3\}$	1.11	3,43
$\{1,4\}$	0.52	3,60
$\{3,4\}$	.42	4.50
$\{1,3,4\}$	2.1	7.34
$\{1,3,4,21\}$	1.49	25.24
$\{1,2,3,4\}$	7.98	9.96
$\{1,2,3,4,21\}$	3.13	24.38

Table 8. Cook's D and F Values for the Stack-Loss Data.

Case	Eigenvalue Coefficients		D
	$\lambda_1 = 2.36$	$\lambda_2 = 1.63$	
1	$-.046$	$-.100$	.040
2	$-.076$	$-.108$	.053
3	$-.016$	$-.118$	.046
4	$-.036$	$-.090$	.031
5	$-.040$	$-.105$	.039
6	$-.053$	$-.103$	.052
7	$-.092$	$-.121$	.079
8	$-.044$	$-.121$	.052
9	$-.030$	$-.098$	.034
10	$-.020$	$-.115$	.047
11	.15	.297	.035
12	$-.01$	.520	.851
13	.24	.149	.254
14	.87	$-.138$	2.11
rest	$\|v_i\| < .032$	$\|v_i\| < .022$	$D_i < .10$

Table 9. Two Largest Eigenvalues and its Eigenvectors and Univariate Cook's D for the Hawkins, Bradu and Kass Data.

set	D	F
$\{11, 13, 14\}$	11.03	28.59
$\{1-10\}$	33.74	109.69
$\{1-10, 14\}$	33.18	98.79
$\{11, 12, 13\}$	4.97	58.45
$\{11, 12, 13, 14\}$	13.37	181.11
$\{1-10, 11, 12, 13\}$	37.42	82.79
$\{1-10, 11, 12, 13, 14\}$	60.18	76.61
$\{11, 12, 13, 14 \mid \{1-10\}\}$	24.33	0.63
$\{1-10 \mid \{11, 12, 13, 14\}\}$	834.89	3.86

Table 10. Cook's D and F Values for the Hawkins, Bradu and Kass Data.

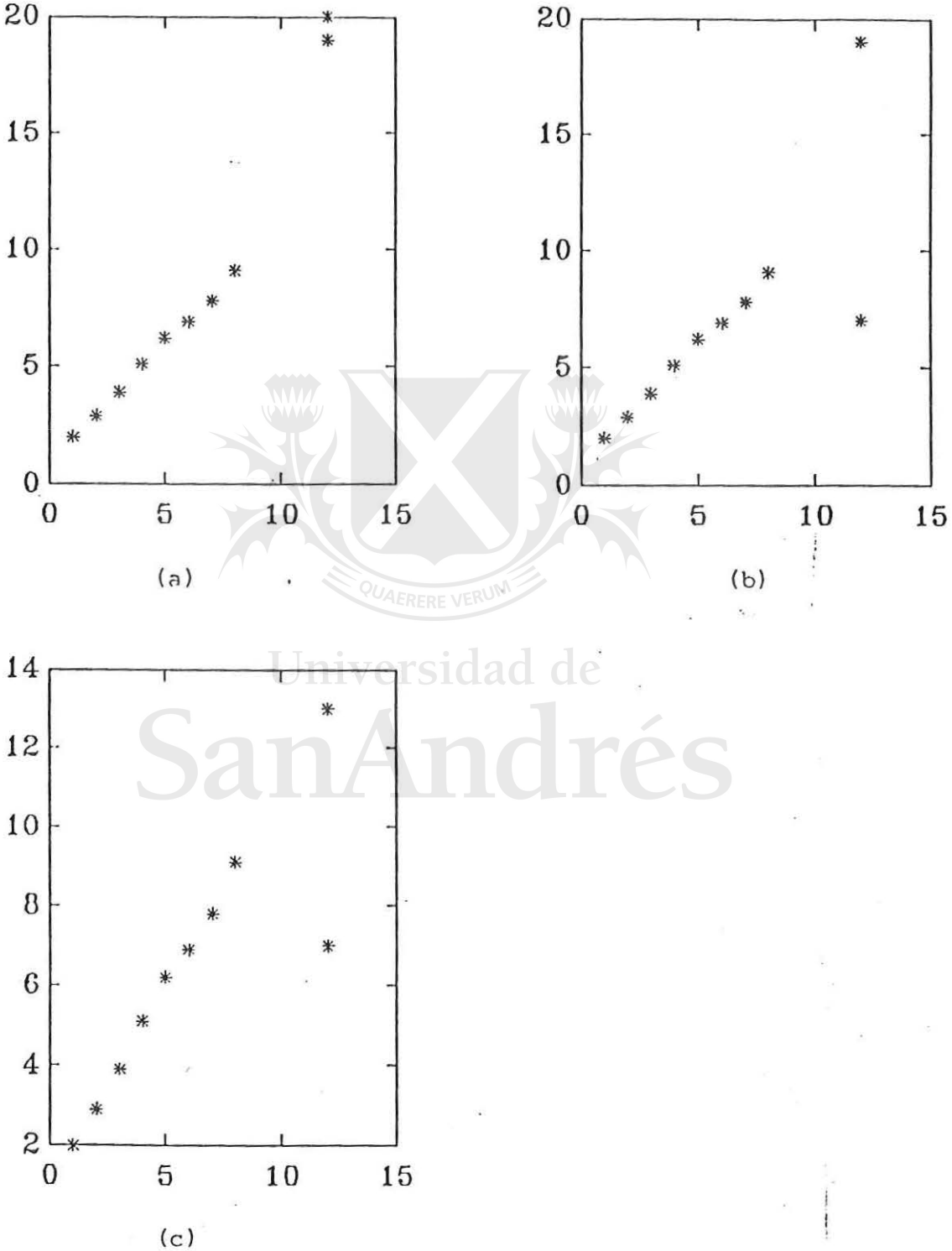


Figure 1. Data for Example 1. The Values are in Table 4.1.

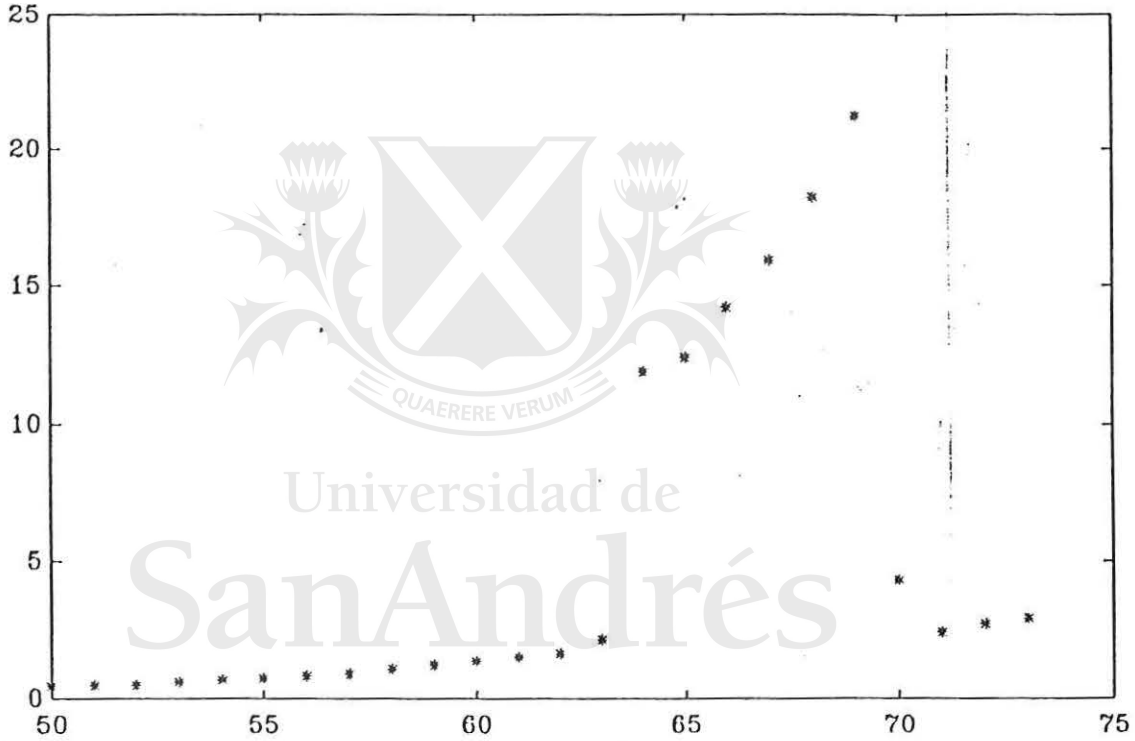


Figure 2. Data of International Phone Calls in Belgium

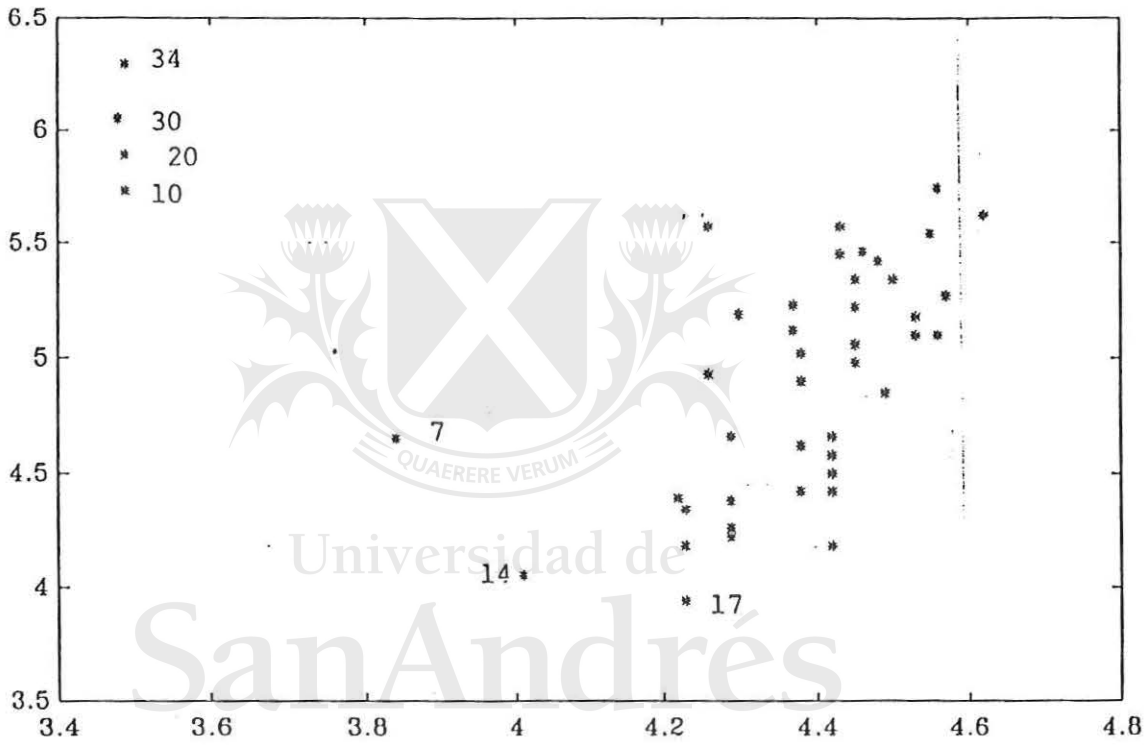
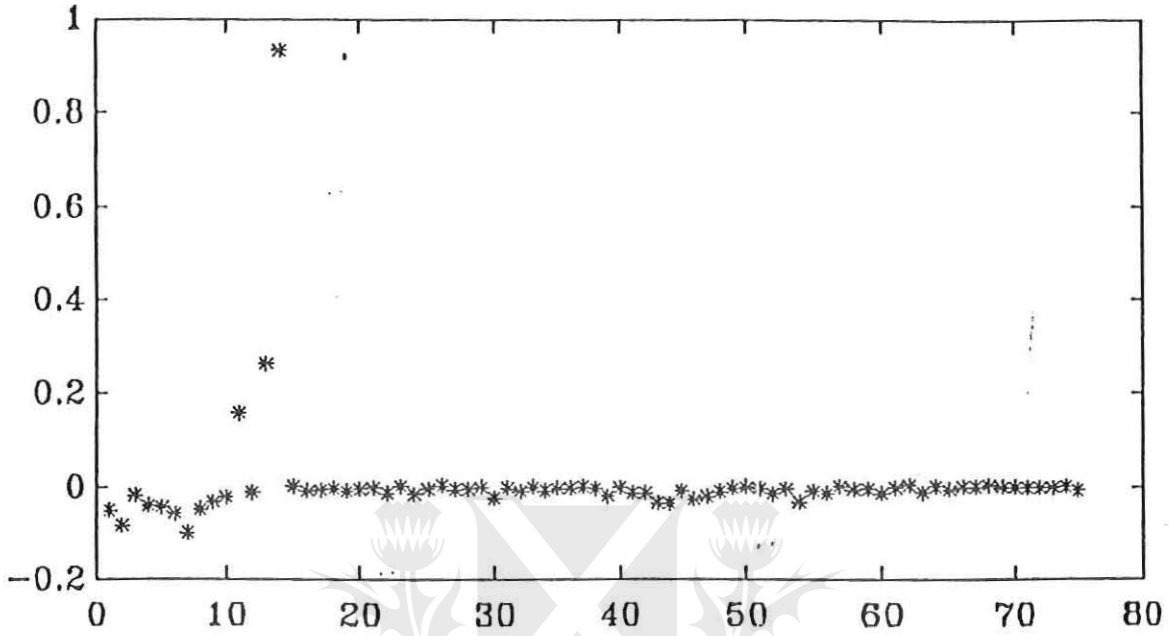


Figure 3. Data of the Hertzsprung-Rusell Diagram of a Star Cluster



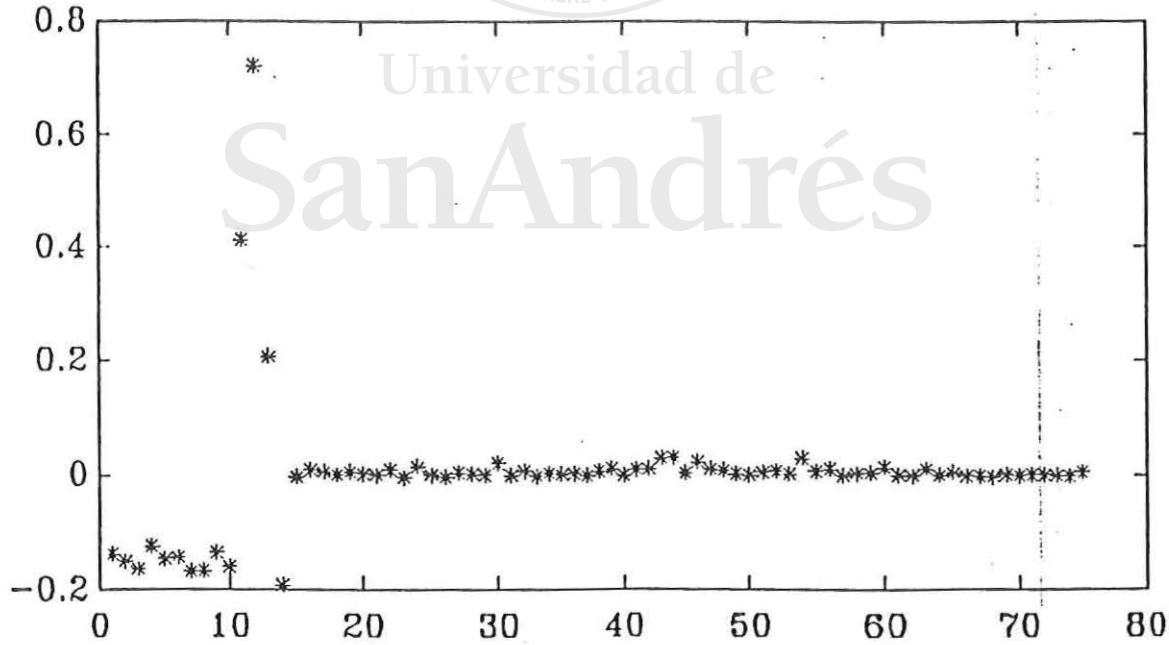


Figure 5. Plot of the Components of First and Second Eigenvectors for the Influence Matrix of Hawkins, Bradu and Kass Data.

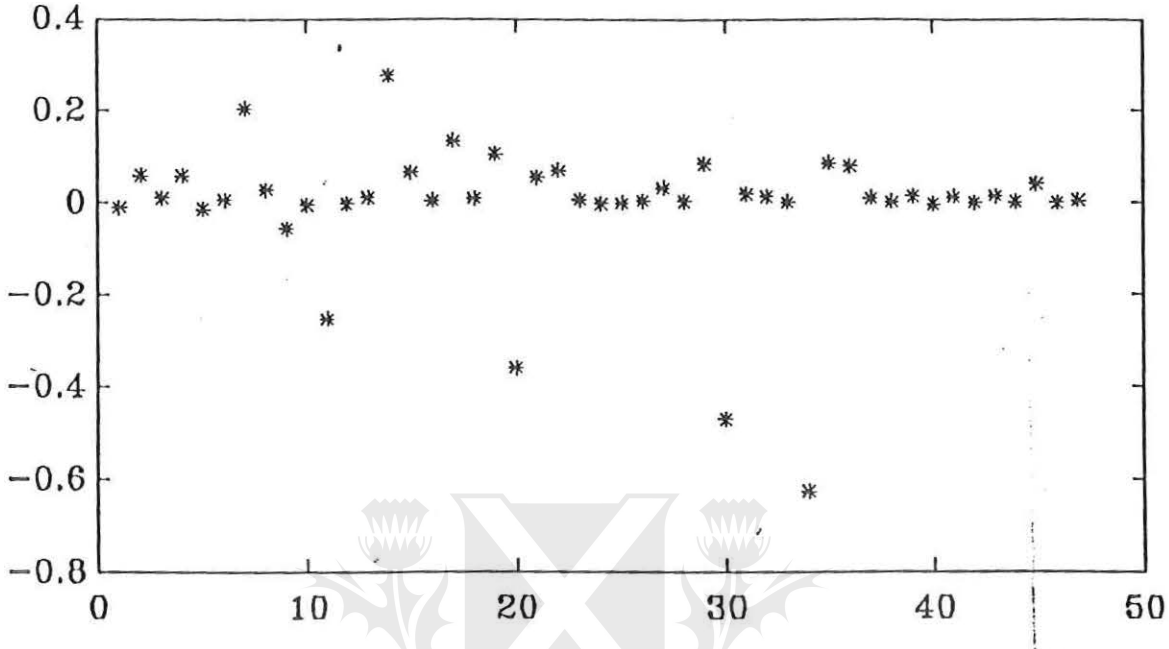


Figure 4. Plot of the Components of First Eigenvector from Table 5.