

Robust Principal Component Analysis with Modified One-Step MEstimator Method

by

FILE	SRIWIJAYANTI_2019_J._PHYS.__CONF._SER._1341_092008.PDF (861.83K)		
TIME SUBMITTED	28-APR-2020 10:55AM (UTC+0700)	WORD COUNT	3209
SUBMISSION ID	1309877474	CHARACTER COUNT	16109

PAPER • OPEN ACCESS

Robust Principal Component Analysis with Modified One-Step M-Estimator Method

6

To cite this article: Sriwijayanti *et al* 2019 *J. Phys.: Conf. Ser.* **1341** 092008

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices
to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of
every title for free.

Robust Principal Component Analysis with Modified One-Step M-Estimator Method

Sriwijayanti¹, Raupong¹ and N Sunusi¹

¹Department of Statistics, Faculty of Mathematics and Natural Sciences,
Hasanuddin University

E-mail: wijyantisri29@gmail.com, Raupong.stat.uh@gmail.com, and
ntitisanusi@gmail.com

Abstract. Principal component analysis (PCA) is a multiple variable analysis method that aims to reduce the dimensions of the original variable, which are mostly correlated so that new variables that are not correlated are obtained. The data used is criminality data in Indonesia in 2016 which contains outlier data in it. Therefore this study cannot use classic PCA because classic PCA was formed based on a covariant variant matrix that is very sensitive to the existence of outlier data. To overcome this problem, PCA robust is used with the Modified One-Step M-Estimator method with a MADn scale estimator to get the main components that are not much influenced by outliers. Modified One-Step M-Estimator (MOM) is the average remaining value of all extreme values that have been issued. The results obtained are there are 3 main components that can explain 85.19% of the variance of the 7 original variables.

1. Introduction

Principal component analysis (PCA) is formed based on the variance covariance matrix which is very sensitive to the existence of outlier data. Therefore we need a PCA model that can handle outlier data. The purpose of Robust PCA is to get the main components that are not much affected by outliers. The first approach is done by replacing the classic variance covariance matrix with a robust estimator [2]. PCA is a multivariate analysis that is used to reduce the dimensions of large and correlated data to be smaller and not correlated with each other but still maintains a minimum data variance of 80% [3].

One method that can be used to obtain robust estimators is Modified One-Step M-Estimator (MOM). MOM is the average remaining value of all extreme values that have been issued. MOM uses Median Absolute Deviation (MADn) as a standard scale estimator. MADn is a robust measures of central tendency and has been claimed in the literature as the best estimator scale on data containing outliers [4]. Based on the description of the background, the author took the title "Robust Principal Component Analysis with Modified One-Step M-Estimator".

2. Materials and Methods

2.1. Eigen Value and Vector Value

Eigenvalue is a value that indicates the influence a variable has on the formation of the characteristics of a vector or matrix. If \mathbf{A} is a matrix of size $n \times n$, then the vector non-trivial \mathbf{x} at \mathbb{R}^n is called the eigenvector of \mathbf{A} . To get an eigen vector from matrix \mathbf{A} of size $n \times n$ it can be written as:

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0} \quad (1)$$

Equation (1) has a non-trivial solution if it satisfies the following equation:

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0 \quad (2)$$

Equation (2) is called the characteristic matrix \mathbf{A} . The scalars that satisfy this equation are the eigenvalues of \mathbf{A} [1].

2.2. Outliers

According to Johnson & Wichern [3], outliers are one or several unusual observations, which appear not to be included in the pattern of variability produced by other observations. Outliers in the data can cause bias in mean and covariance estimators of data multivariate [9]. One method that can be used to detect outliers in multivariate data is the Mahalanobis distance. Mahalanobis distance is obtained by calculating the distance of each observation of the data center. The use of Mahalanobis distance is not maximal if the data contains more than one outlier observation. This arises due to the influence of masking and swamping. The effect of masking occurs when outliers are not detected as outliers because there are other outliers observations nearby. The effect of swamping occurs when observations not outliers are identified as outliers observations. The effect of masking and swamping can be overcome with robust estimators [7].

One robust estimator is MOM-MADn to estimate the mean vector and covariant variance matrix used to estimate the Mahalanobis distance. The i^{th} observation is defined as multivariate outlier data if the Mahalanobis distance ($d_{i,\text{MOM-MADn}}^2$) is greater than the chi-square table value ($\chi_{\alpha,p}^2$) on the variable p . The mahalanobis distance square is calculated by the formula [3] as follows:

$$d_{i,\text{MOM-MADn}}^2 = (\mathbf{x}_i - \boldsymbol{\mu}_{\text{MOM-MADn}})^T \boldsymbol{\Sigma}_{\text{MOM-MADn}}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_{\text{MOM-MADn}}) > \chi_{\alpha,p}^2 \quad (3)$$

With $\boldsymbol{\mu}_{\text{MOM-MADn}}$ and $\boldsymbol{\Sigma}_{\text{MOM-MADn}}$ the average vector and the covariant variant matrix are estimated using the MOM-MADn method.

2.3. Principal Component Analysis

PCA is a multiple variable analysis methods that aims to reduce the dimensions of the original variable which are mostly correlated so that few variables (principal component) are not correlated with each other, where the component is a linear combination of original variables such that it has maximum variance [4]. The principal component formed can be written as follows:

$$\begin{aligned} K_1 &= \mathbf{v}_1^T \mathbf{Z} = v_{11}Z_1 + v_{12}Z_2 + \dots + v_{1p}Z_p \\ K_2 &= \mathbf{v}_2^T \mathbf{Z} = v_{21}Z_1 + v_{22}Z_2 + \dots + v_{2p}Z_p \\ &\vdots \\ K_p &= \mathbf{v}_p^T \mathbf{Z} = v_{p1}Z_1 + v_{p2}Z_2 + \dots + v_{pp}Z_p \end{aligned}$$

or

$$\mathbf{K} = \begin{bmatrix} K_1 \\ K_2 \\ \vdots \\ K_p \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1p} \\ v_{21} & v_{22} & \dots & v_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ v_{p1} & v_{p2} & \dots & v_{pp} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix} = \mathbf{VZ} \quad (4)$$

with:

1. $\text{var}(K_j) = \text{var}(\mathbf{v}_j^T \mathbf{Z}) = \mathbf{v}_j^T \text{var}(\mathbf{Z}) \mathbf{v}_j$
so that
 $\text{var}(K_j) = \mathbf{v}_j^T \boldsymbol{\Sigma} \mathbf{v}_j ; j = 1, 2, \dots, p$
with $\boldsymbol{\Sigma} = \text{var}(\mathbf{Z}) = \boldsymbol{\Sigma}_Z$
2. $\text{cov}(K_j, K_i) = \text{cov}(\mathbf{v}_j^T \mathbf{Z}, \mathbf{v}_i^T \mathbf{Z}) = \mathbf{v}_j^T \text{cov}(\mathbf{Z}, \mathbf{Z}) \mathbf{v}_i = \mathbf{v}_j^T \boldsymbol{\Sigma} \mathbf{v}_i$
So that

$$\begin{aligned} \text{cov}(K_j, K_l) &= v_j^T \Sigma v_l ; \quad j, l = 1, 2, \dots, p \\ \text{with } \text{cov}(Z, Z) &= \text{var}(Z) = \Sigma_Z \end{aligned}$$

2.4. Median Absolute Deviation

MAD_n is often used as an initial value for a more efficient robust estimator calculation. MAD_n for random samples n observation X_1, X_2, \dots, X_n is defined as follows [6]:

$$MAD_n = 1.4826 \text{Med} \{ |X_{ij} - \text{Med}_{X_j}| \} , \quad i = 1, 2, \dots, n \quad (5)$$

with Med is the sample median.

2.5. Modified One-Step M-Estimator

MOM is the average remaining value of all extreme values that have been issued. This estimator is derived from the one-step M-estimator Staudte & Sheather [8] after some modification. Mathematically, Wilcox [10] defines the MOM estimator as

$$\hat{\theta}_j = \frac{\sum_{i=i_1+1}^{n_j-i_2} X_{(i)j}}{n_j - i_1 - i_2} \quad (6)$$

where:

- $X_{(i)j}$ = is the i^{th} order observation in the j^{th} characteristic variable
- i_1 = is the number of $X_{(i)j}$ that satisfies the criteria $(X_{(i)j} - \hat{M}_j) < -K \times MAD_n$
- i_2 = is the number of $X_{(i)j}$ that satisfies the criteria $(X_{(i)j} - \hat{M}_j) > K \times MAD_n$
- n_j = is size of the data set for each variable
- \hat{M}_j = is the median of the data in each j^{th} variable

and the scale estimator is the median absolute deviation MAD_n . The constant $K = 2.24$ is motivated to give a good efficiency for the robust scale estimators MAD_n when the sample is taken from a normal distribution [5].

3. Analysis Method

3.1. Data Source

In this study the data used were simulation data and secondary data on criminality data in Indonesia in 2016 obtained from the 2018 Criminal Statistics book of the Central Statistics Agency (BPS) of the Republic of Indonesia. Criminal data consists of variables calculated from the number of crime cases that occur per 100,000 population consisting of 31 observations and 7 variables.

3.2. Description of Variable

The variables used in this study are:

- X_1 : Number of murder cases
- X_2 : Number of cases of abuse
- X_3 : Number of rape cases
- X_4 : Number of theft cases
- X_5 : Number of cases of damage
- X_6 : Number of cases of narcotics abuse
- X_7 : Number of fraud cases

3.3. Analytical Method

3.3.1. Simulation Methods

The simulation data used is 100 samples and 7 variables. Following are the analysis steps for simulation data:

1. Settings parameters \bar{x} , \bar{x}^p , dan Σ , with value:

$$\bar{x} = [96 \quad 105 \quad 97 \quad 112 \quad 100 \quad 109 \quad 91]$$

$$\bar{x}^p = [155 \quad 163 \quad 172 \quad 148 \quad 169 \quad 157 \quad 175]$$

$$\Sigma = \begin{bmatrix} 116 & -35 & 77 & 41 & 40 & 25 & 8 \\ -35 & 165 & -2 & 30 & 16 & 59 & 24 \\ 77 & -2 & 302 & -31 & 58 & 66 & 67 \\ 41 & 30 & -31 & 140 & 6 & 113 & 41 \\ 40 & 16 & 58 & 6 & 247 & 32 & 82 \\ 25 & 59 & 66 & 113 & 32 & 218 & 171 \\ 8 & 24 & 67 & 41 & 82 & 171 & 184 \end{bmatrix}$$

With \bar{x} is the average vector for non-outliers, \bar{x}^p is the average vector for outliers and Σ is the variance covariance matrix.

2. Generating multivariate normal data $X \sim MN(\bar{x}, \Sigma)$ with $n_1 = 95$.
3. Generating multivariate normal data for outliers of 5% $X^{p1} \sim MN(\bar{x}^p, \Sigma)$ with the number of outliers $n_1^* = 5$.
4. Combine the data in point 2 and the data in point 3 that has been raised.
5. Standardize simulation data.
6. Calculate eigenvalues and eigenvectors based on the MOM-MADn method.
7. Determine the number of main components selected based on the percentage of cumulative variance proportions from the MOM-MADn method.
8. Form the principal component based on the formula in Equation (4).
9. Repeat steps two to eight with $n_2 = 90$ and outliers of $n_2^* = 10$.

3.3.2. Methods for Secondary Data

1. Calculate the correlation matrix in the data.
2. Estimating the average vector, the covariant variant matrix and the correlation matrix with the MOM-MADn method with the following algorithm:
 - a. Calculate MADn by usng Equation (5).
 - b. Crimming data outliers on MOM criteria using formulas $|X_{ij} - \hat{M}_j| > K \times (MAD_n)$.
 - c. Calculate $\hat{\theta}_j = (x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6 \quad x_7)$ by using Equation (6).
 - d. Calculate the mean vector $\bar{x}_{MOM-MADn}$ based on points (c).
 - e. Calculate $\Sigma_{MOM-MADn}$ MOM-MADn variance covariance matrix using $\bar{x}_{MOM-MADn}$.
 - f. Calculate the correlation matrix of the MOM-MADn variance covariance matrix.
3. Calculate the mahalanobis distance for each data, using the formula:

$$d_{i,MOM-MADn}^2 = (x_i - \mu_{MOM-MADn})^t \Sigma_{MOM-MADn}^{-1} (x_i - \mu_{MOM-MADn})$$

4. Detect outliers by comparing $d_{i,MOM-MADn}^2$ with the value $\chi_{0.05,7}^2 = 14,06714$.
5. Standardize criminal data.
6. Calculates eigen values and eigen vectors based on the correlation matrix of the MOM-MADn method using the formula:

$$\det(\rho - \lambda I) = 0$$

$$(\rho - \lambda I)x = 0$$

7. Determine the number of principal component selected based on the percentage of cumulative variance proportions from the MOM-MADn method.
8. Form the principal component based on the formula in Equation (4).
9. Interpretation of data

4. Result and Discussion

4.1. Robust Principal Component Analysis Simulation Data

Robust PCA MOM-MADn starting with estimating parameters, namely by replacing \bar{x} and S with $\bar{x}_{\text{MOM-MADn}}$ and $\Sigma_{\text{MOM-MADn}}$ which is the average vector and variance covariance matrix with the MOM-MADn method. Before estimating the data parameters, it must be standardized first.

4.1.1. Calculate Eigenvalue and Eigenvector

Based on Equation (2) the eigenvalues obtained for simulation data are as follows:

Table 4.1 Eigenvalue Simulation Data

Parameter	Outliers 5 %	Outliers 10%
λ_1	4,9976	5,8516
λ_2	0,7091	0,5891
λ_3	0,5726	0,3238
λ_4	0,4179	0,3134
λ_5	0,3570	0,2025
λ_6	0,1163	0,1004
λ_7	0,0259	0,0179

Based on the formula in Equation (1), the normalized eigenvectors are obtained as follows:

Table 4.2 Eigen vector Simulation Data

	v_1	v_2	v_3	v_4	v_5	v_6	v_7
Outliers 5%	0,3799	-0,2583	-0,5596	0,2771	0,1902	0,5717	0,1897
	0,3618	-0,0192	0,5642	-0,0003	0,7390	0,0407	0,0508
	0,3785	-0,3248	-0,3955	-0,5128	0,1488	-0,5298	-0,1705
	0,3388	0,6770	-0,2439	0,4345	0,0746	-0,2968	-0,2913
	0,3674	-0,4273	0,2744	0,5119	-0,3966	-0,3632	0,2364
	0,3927	0,4306	0,0670	-0,4108	-0,2805	0,0812	0,6339
Outliers 10%	0,4213	-0,0318	0,2697	-0,2018	-0,3925	0,4052	-0,6243
	0,3935	-0,0455	-0,4625	0,3789	-0,3809	-0,5099	-0,2837
	0,3739	0,1352	0,4224	-0,5723	-0,5709	-0,0617	-0,0786
	0,3818	0,3200	0,1969	0,5776	-0,1959	0,5421	0,2173
	0,3267	-0,7739	-0,2517	-0,1514	-0,0815	0,3265	0,3081
	0,3611	0,4329	-0,5501	-0,3975	0,3218	0,2980	-0,1701
	0,3875	-0,2773	0,4210	0,1183	0,4621	0,0219	-0,6064
0,4151	0,1182	0,1603	-0,0204	0,4084	-0,4963	0,6119	

The eigenvectors in Table 4.2 are eigen vectors, each of which corresponds to the eigen values in Table 4.1, meaning v_j is obtained λ_j , where $j = 1, 2, \dots, 7$.

11

4.1.2. Determine the Number of Principal Component

Determination of the number of principal component (PC) formed is based on the percentage of the cumulative variance proportion obtained at least 80% obtained from eigenvalues.

Table 4.3 Proportion of Cumulative Variance and Variance of Simulation Data

Principal Components	Outliers 5 %		Outliers 10%	
	Proporsion Variance	Proporsion Variance Kumulatif	Proporsion Variance	Proporsion Variance Kumulatif

1	69,44%	69,44%	79,09%	79,09%
2	9,85%	79,29%	7,96%	87,05%
3	7,96%	87,25%	4,38%	91,43%
4	5,81%	93,06%	4,24%	95,67%
5	4,96%	98,02%	2,74%	98,40%
6	1,62%	99,64%	1,36%	99,76%
7	0,36%	100%	0,24%	100%

Table 4.2 shows that data with outliers of 5% and 10% in the first PC each explained 69.44% and 79.09% of the total sample variance. The first and second PC cumulatively each explained 79.29% and 87.05% of the total sample variance. The first, second, and third PC cumulatively have each explained 87.25% and 91.43% of the total sample variance.

4.1.3. Form the Principal Component

Based on the cumulative variance proportions in Table 4.3 it is obtained for simulation data with outliers of 5% and 10% respectively, enough use 3 PC and 2 PC able to explain 87.25% and 87.05% variance of the origin variables.

1. Simulation data with outliers of 5%

To form 3 PC, 3 eigenvectors are needed which correspond to λ_1 , λ_2 , and λ_3 namely v_1 , v_2 , and v_3 . Based on Equation 2.11 3 PC will be formed which are linear combinations of the origin variables as follows:

$$K_p = v_p^t Z = v_{p1} Z_1 + v_{p2} Z_2 + \dots + v_{pp} Z_p, \quad p = 1, 2, \dots, 7$$

$$K_1 = v_1^t Z = 0,3799 Z_1 + 0,3618 Z_2 + 0,3785 Z_3 + 0,3388 Z_4 + 0,3674 Z_5 + 0,3927 Z_6 + 0,4213 Z_7$$

$$K_2 = v_2^t Z = -0,2583 Z_1 - 0,0192 Z_2 - 0,3248 Z_3 + 0,6770 Z_4 - 0,4273 Z_5 + 0,4306 Z_6 - 0,0318 Z_7$$

$$K_3 = v_3^t Z = -0,5596 Z_1 + 0,5647 Z_2 - 0,3955 Z_3 - 0,7439 Z_4 + 0,2744 Z_5 + 0,0670 Z_6 + 0,2697 Z_7$$

2. Simulation data with outliers of 10%

To form 2 PC, 2 eigenvectors are needed which correspond to λ_1 and λ_2 namely v_1 and v_2 . Based on Equation 2.11 2 PC will be formed which are linear combinations of the origin variables as follows:-

$$K_p = v_p^t Z = v_{p1} Z_1 + v_{p2} Z_2 + \dots + v_{pp} Z_p, \quad p = 1, 2, \dots, 7$$

$$K_1 = v_1^t Z = 0,3935 Z_1 + 0,3739 Z_2 + 0,3818 Z_3 + 0,3267 Z_4 + 0,3611 Z_5 + 0,3875 Z_6 + 0,4151 Z_7$$

$$K_2 = v_2^t Z = -0,0455 Z_1 + 0,1352 Z_2 + 0,3200 Z_3 - 0,7739 Z_4 + 0,4329 Z_5 - 0,2773 Z_6 + 0,1182 Z_7$$

4.2. Robust Principal Component Analysis Secondary Data

Robust PCA with the MOM method uses the MADn estimator scale conducted on criminality data in Indonesia in 2016. The first step is to estimate parameters on robust PCA by replacing x , Σ , and ρ with $\bar{x}_{MOM-MADn}$, $\Sigma_{MOM-MADn}$, and $\rho_{MOM-MADn}$. Before estimating parameters the data it must be standardized first.

4.2.1. Outliers Detection

If mahalanobis distance ($d_{i,MOM-MADn}^2$) > 14,06714 the data will be said as outlier data..

Table 4.4 Mahalanobis Distance from data identified as outliers

Observation	Province	$d_{i,MOM-MADn}^2$
1	North Sumatra	24,022
2	Riau	18,059

3	South Sumatra	26,651
4	Metro Jaya	24,671
5	East Java	23,859

Based on the detection process, 5 out of 31 observational data were identified as outliers observations. The provinces identified as outliers observations are North Sumatra, Riau, South Sumatra, Metro Jaya, and East Java.

4.2.2. Calculate Eigenvalue and Eigenvector

Based on the formula in Equation (2) the eigenvalue is obtained as follows:

Table 4.5 Criminal Data Eigenvalue

Parameter	Eigen Value
λ_1	4,3663
λ_2	0,9626
λ_3	0,6343
λ_4	0,4530
λ_5	0,2439
λ_6	0,2012
λ_7	0,1387

Based on the formula in Equation (1), the normalized eigenvectors are obtained as follows:

Table 4.6 Criminal Data Eigenvector

v_1	v_2	v_3	v_4	v_5	v_6	v_7
-0,1631	-0,7344	0,3816	-0,2118	0,1340	0,2111	0,4254
-0,2369	-0,0169	-0,8225	-0,1431	0,2161	0,0759	0,4407
0,3272	0,0124	0,0234	0,8223	0,1555	0,2334	0,3706
-0,3276	0,0171	0,0020	0,1819	-0,8551	-0,1377	0,3304
0,1513	0,5476	0,2266	-0,3814	-0,0908	0,5952	0,3434
-0,4165	0,3960	0,3529	0,0553	0,3998	-0,5035	0,3588
0,7117	-0,0564	-0,0378	-0,2771	-0,1092	-0,5181	0,3628

The eigenvectors in Table 4.6 are eigenvectors which correspond to the eigenvalues in Table 4.5, meaning v_j is obtained from λ_j , where $j = 1, 2, \dots, 7$.

4.2.3. Determine Number of Principal Components

Determination of the number of PCs formed is based on the percentage proportion of cumulative variance obtained at least 80% from eigenvalues.

Table 4.7 Proportion of Variances and Cumulative Variance of Criminal Data

Principal Component	variance proportion	cumulative variance proportion
1	62,38%	62,38%
2	13,75%	76,13%
3	9,06%	85,19%
4	6,47%	91,66%
5	3,49%	95,15%
6	2,87%	98,02%
7	1,98%	100%

Table 4.7 shows that on the first PC explained 62.38% of the total sample variance. The first and second PC cumulatively explained 76.13% of the total sample variance. The first, second, and third PC cumulatively explained 85.19% of the total sample variance.

4.2.4. Form the Principal Component

Based on the cumulative variance proportion obtained 3 PC which is able to explain 85.19% the variance of the origin variables. Therefore to form PC are needed 3 eigenvectors which correspond to λ_1 , λ_2 , and λ_3 , namely v_1 , v_2 , and v_3 .

$$K_p = v_p^t Z = v_{p1} Z_1 + v_{p2} Z_2 + \dots + v_{pp} Z_p, \quad p = 1, 2, \dots, 7$$
$$K_1 = v_1^t Z = -0,1631 Z_1 - 0,2369 Z_2 + 0,3272 Z_3 - 0,3276 Z_4 + 0,1513 Z_5 - 0,4165 Z_6 + 0,7117 Z_7$$
$$K_2 = v_2^t Z = -0,7344 Z_1 - 0,0169 Z_2 + 0,0124 Z_3 + 0,0171 Z_4 + 0,5476 Z_5 + 0,3960 Z_6 - 0,0564 Z_7$$
$$K_3 = v_3^t Z = 0,3816 Z_1 - 0,8225 Z_2 + 0,0234 Z_3 + 0,0020 Z_4 + 0,2266 Z_5 + 0,3529 Z_6 - 0,0378 Z_7$$

Based on the PC that has been produced, the principal component score will be calculated. The resulting PC score is new variables from PCA results which are linear combinations of the origin variables.

5. Conclusion

Based on the results and discussion, conclusions are obtained as follows:

1. Based on the Modified One-Step M-Estimator method with the MADn estimator scale at criminality data in Indonesia 2016 obtained 5 of 31 observational data that have mahalanobis distance large than chi-square table values ($\hat{d}_{i,MOM-MADn}^2 > 14,06714$) identified as outliers observations, namely: North Sumatra, Riau, South Sumatra, Metro Jaya, and East Java provinces.
2. Reduction of data dimensions using the modified One-Step M-Estimator method with the MADn estimator scale in criminality data in Indonesia in 2016 resulted in 3 principal components which were able to explain 85.19% of the variance of the 7 origin variables, namely: murder cases, persecution cases, rape cases, theft cases, destruction cases, narcotics abuse cases, and fraud cases.

References

- [1] Anton, H., & Rorres, C. 2005. Elementary Linear Algebra, Applications Version, Ninth Edition: John Wiley & Sons.
- [2] Hubert M, Rousseeuw PJ, Vandenberghe K. 2005. *ROBPCA: A New Approach to Robust Principal Component Analysis*. *Technometrics*. 47: 64-79. Another reference
- [3] Johnson, R.A. & Wichern, D.W. 2007. *Applied Multivariate Statistical Analysis*, Sixth Edition. New Jersey : United States of America.
- [4] Melik, H.N. 2017. *Robust Linear Discriminant Analysis Using MOM- Q_n And WMOM- Q_n Estimators: Coordinate-Wise Approach*. Malaysia : University Utara Malaysia.
- [5] Othman, A. R., Keselman, H. J., Padmanabhan, A. R., Wilcox, R. R., & Fradette, K. 2004. Comparing measures of the " typical " score across treatment groups. *British Journal of Mathematical and Statistical Psychology*. 57, 215 - 234.
- [6] Rousseeuw, P.J. and C. Croux, 1993. *Alternatives to the median absolute deviation*. *J. Am. Stat. Assoc.*, 80: 1273-1283.
- [7] Rousseeuw, P.J., & Van Zomeren, B.C. 1990. Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*. 85(411), 633-651.
- [8] Staudte, R. G., & Sheather, S. J. 1990. *Robust estimation and testing: Wiley series in probability and mathematical statistics*. New York.

- [9] Wang, M., Martin R., Mao, G. 2015. *A Nonsingular Robust Covariance In Multivariate Outlier Detection*. Wilrijk, Belgium: Department of Mathematics and Computer Science, University of Antwerp (UIA).
- [10] Wilcox, R., 2002. Multiple comparisons among dependent groups based on a modified one-step M-estimator. *Journal of Applied Statistics.*, 30(10), 1231-1241.

Robust Principal Component Analysis with Modified One-Step MEstimator Method

ORIGINALITY REPORT

% **10**
SIMILARITY INDEX

% **4**
INTERNET SOURCES

% **7**
PUBLICATIONS

% **6**
STUDENT PAPERS

PRIMARY SOURCES

1 www.ukessays.com % **1**
Internet Source

2 Albassam, Tameem(AI Shawi). "A model for measuring service quality in internet-based services: The case study of internet banking services in the Kingdom of Saudi Arabia", Brunel University School of Engineering and Design PhD Theses, 2013. % **1**
Publication

3 Submitted to Universiti Utara Malaysia % **1**
Student Paper

4 Roongrasamee Boondao. "Crime Risk Factors Analysis", Statistics in Practice, 03/03/2008 % **1**
Publication

5 Hubert, Mia. "Robust Calibration", Practical Guide To Chemometrics Second Edition, 2006. % **1**
Publication

6 A. F. Ozdemir, Rand R. Wilcox, E. Yildiztepe. "Comparing Measures of Location: Some Small- % **1**

Sample Results When Distributions Differ in Skewness and Kurtosis Under Heterogeneity of Variances", Communications in Statistics - Simulation and Computation, 2012

Publication

7	doc.lagout.org Internet Source	% 1
8	Nora Muda. "The Phylogenetic Tree of RNA Polymerase Constructed Using MOM Method", 2009 International Conference of Soft Computing and Pattern Recognition, 12/2009 Publication	% 1
9	www.scipub.org Internet Source	<% 1
10	epdf.tips Internet Source	<% 1
11	Submitted to Universiti Sains Malaysia Student Paper	<% 1
12	Kotas, M.. "Application of projection pursuit based robust principal component analysis to ECG enhancement", Biomedical Signal Processing and Control, 200610 Publication	<% 1
13	Submitted to Kerbala University Student Paper	<% 1

14

Submitted to University of South Africa

Student Paper

<% 1

15

epdf.pub

Internet Source

<% 1

16

www.stat-d.si

Internet Source

<% 1

17

Peter J. Rousseeuw, Michiel Debruyne, Sanne Engelen, Mia Hubert. "Robustness and Outlier Detection in Chemometrics", Critical Reviews in Analytical Chemistry, 2006

Publication

<% 1

18

www.slideshare.net

Internet Source

<% 1

19

lagrange.math.siu.edu

Internet Source

<% 1

20

Submitted to Associatie K.U.Leuven

Student Paper

<% 1

EXCLUDE QUOTES ON

EXCLUDE
BIBLIOGRAPHY ON

EXCLUDE MATCHES

< 5
WORDS