

Factor analysis based on classical and robust estimators

Muthukrishnan R¹, E D Boobalan^{2*}

^{1,2}Department of Statistics, Bharathiar University, Coimbatore-46 Tamil Nadu, INDIA.

Email: muthukrishnan70@rediffmail.com, boobalanstat@gmail.com

Abstract

Introduction: The multivariate methods, such as principal component analysis, discriminant analysis, cluster analysis, multivariate regressions etc., are mainly based on the empirical measures mean vector, covariance and correlation matrices. All these measures strongly affected by even a single outliers present in the multivariate data set. Robust alternatives measures are established to overcome this limitation. Many multivariate robust procedures are established to estimate these measures. All these robust procedures established based on the sample of selecting the best observations (which represents the original data) nearly half of the data points. Among these, the minimum covariance determinant estimator (MCD) proposed by Rousseeuw (1984) is one of the highly robust estimators of estimating multivariate location and scatter. This paper provides an attempt to explore such robust procedures along with the application in factor analysis. Further it is proposed to construct robust factor analysis with the help of most widely used robust methods MVE, S and MM that can resist the effect of outliers. The efficiency of these estimators with classical one is carried out by providing an empirical study with a help of MATLAB software.

Keywords: Multivariate Methods, Robust Estimators, MATLAB, Simulation Study, Factor Loadings.

*Address for Correspondence:

Dr. E. D. Boobalan, Department of Statistics, Bharathiar University, Coimbatore-46. Tamil Nadu, INDIA.

Email: boobalanstat@gmail.com

Received Date: 14/08/2014 Accepted Date: 24/08/2014

Access this article online

Quick Response Code:



Website:

www.statperson.com

DOI: 03 September
2014

INTRODUCTION

Multivariate analysis is a statistical technique for simultaneous analysis of two or more variables observed from one or more sample objects. The objective of the analysis is to estimate the extend or amount of relationship among the variables. When working with p-dimensional multivariate normal data both the location and scatter are of interest. The location is described by a mean vector which represents a point in the multidimensional space and the scatter is described by a variance-covariance matrix. The sample mean vector and the sample covariance matrix are the corner stone of the classical multivariate analysis. They are optimal when the underlying data are normal. They, however, are notorious for being extremely sensitive to outliers and heavy tailed data. Robust alternatives of these classical location and scatter estimators are available. These types of estimators indeed are much more robust against outliers and contaminated data. This paper provides a brief description on the robust estimators MCD, MVE, S and MM. It is proposed to construct factor analysis using these robust estimators and efficiencies are measured with classical factor analysis. The brief introduction about factor analysis along with robust and classical counterpart is discussed in section 2. Section 3 provides classical and robust estimators. The performance of the proposed method has been carried out with numerical experiments and the results are provided in the section 4. The findings and discussions are presented in the last section.

CLASSICAL AND ROBUST FACTOR ANALYSIS

Factor analysis is a popular multivariate technique. Its goal is to approximate the p original variables of the dataset by linear combinations of a smaller number k of latent variables, called factors. The classical factor analysis (FA) starts with the usual sample covariance (or correlation) matrix and then the eigenvectors and eigenvalues of the matrix are employed for estimating the loading matrix. This must be done in such a way that the covariance matrix or the correlation matrix of the p original variables is fitted well. The factor analysis model contains many parameters, including the specific variances of the error components. The classical technique starts by computing the usual sample covariance matrix or the sample correlation matrix, followed by a second step which decomposes this matrix according to the model. This approach is not robust to outliers in the data, since they already have a large effect on the first step. The analysis, however, is not robust since outliers can have a large effect on the covariance (or correlation matrix) and the results obtained may be misleading or unreliable. A straightforward approach to robustify the classical FA is to replace the sample covariance (or correlation) matrix with a robust one. Therefore construct a robust factor analysis method, which in the first step computes a highly resistant scatter matrix such as the minimum covariance determinant (MCD) estimator (Rousseeuw 1985, 1999), Rousseeuw's minimum volume ellipsoid (MVE) estimator, Rousseeuw and Yohai's S-estimators and Huber's M-estimators [Campbell (1980, 1982); Davies (1987); Hampel, Ronchetti, Rousseeuw and Stahel (1986); Huber (1981); Kent and Tyler (1991); Lopuhaa (1989); Lopuhaa and Rousseeuw (1991); Maronna (1976); Rousseeuw (1985); Rousseeuw and Leroy (1987); Rousseeuw and Yohai (1984); Rousseeuw and van Zomeren (1990a, 1990b, 1991); Tyler (1983, 1988, 1991)]. For the second step several methods are available, such as maximum likelihood estimation and the principal factor analysis method.

CLASSICAL AND ROBUST ESTIMATORS

Maximum Likelihood Estimator (MLE)

The principle of maximum likelihood estimation (MLE), originally developed by R.A Fisher in 1920. Assuming that the data is drawn from a population whose distribution is multivariate normal, then the optimal estimators for location and dispersion are found, respectively, as the $P \times 1$ sample mean vector,

$$m = \frac{\sum_{i=1}^n Z_{y,i}}{n} \quad (1)$$

and $P \times P$ sample covariance matrix

$$C = \frac{\sum_{i=1}^n (Z_{y,i} - m)(Z_{y,i} - m)'}{n - 1} \quad (2)$$

These are, obviously, mean-based estimators, so any unusual or extreme observation can arbitrarily inflate either of them.

Robust Estimator

The Minimum Volume Ellipsoid (MVE) estimator was first proposed by Rousseeuw (1984). It has been frequently used in detection of multivariate outliers. The estimation seeks to find the ellipsoid of minimum volume that covers a subset of at least h data points. Subsets of approximately 50% of the observations are examined to find the subset that minimizes the volume occupied by the data. The best subset (smallest volume) is then used to calculate the covariance matrix and the Mahalanobis distances to all the data points. An appropriate cut-off value is then estimated, and the observations with distances that exceed that cut-off are declared to be outliers. To minimize computation time, Rousseeuw and Leroy (1987) proposed a resampling algorithm in which subsamples of $p+1$ observations (p is the number of variables), the minimum to determine an ellipsoid in p -dimensional space, are initially drawn. Another robust estimator, minimum covariance determinant estimator (MCD) proposed by Rousseeuw (1984, 1985) is a highly robust estimator of multivariate location and scatter. In beginning of 1984 when Rousseeuw introduced nobody didn't use it due to lack of information about the calculating procedure and also time consuming, so in practice one resort to approximate algorithms. After that the algorithm modified for the computation purpose. To overcome this limitation Rousseeuw (1999) introduced a new algorithm is called FAST-MCD algorithm. It is contain concentration step (C-step) procedure to simplify the computation process. A key step of new algorithm is the fact that starting from any approximation to the MCD, it is possible to compute another approximation with an even lower determinant. The FAST-MCD method is able to handle large data sets within a reasonable amount of time. In fact, Rousseeuw and Van Driessen (1999) successfully

analyzed with large data. Rousseeuw and Yohai (1984) introduced S estimator which is slightly different from the existing robust estimators. Also the authors studied the existence, consistency, asymptotic normality and breakdown point of the estimator. Davies (1987) investigated some properties of S-estimators of multivariate location and covariance. An S-estimator of multivariate location and scale minimizes the determinant of the covariance matrix, subject to a constraint on the magnitudes of the corresponding Mahalanobis distances. The multivariate MM-estimator was introduced by Tatsuoka and Tyler (2000) as belonging to a broad class of estimators namely multivariate M-estimators with auxiliary scale. M-estimator was originally constructed by Huber (1964) for the estimation of a one-dimensional location parameter. Maronna (1976) was the first to define M-estimator for multivariate location and covariance. The idea is to estimate the scale by means of a very robust S-estimator, and then estimate the location and shape using a different ρ -function that yields better efficiency at the central model. The location and shape estimates inherit the breakdown point of the auxiliary scale and can be seen as a generalization of the regression MM-estimators of Yohai (1987).

Numerical Study

This section presents the performance of classical and various robust procedures, particularly MCD, MVE, S and MM are considered for the construction of factor analysis. Factor loadings of each variable by each factor under various procedures along with plots are also discussed in this section. The numerical study is carried out using MATLAB software which includes two packages namely forward Search Data Analysis (FSDA), Library for Robust Analysis (LIBRA). The study also provides results under different level of contamination of data.

Experiment 1

The factor analysis has performed in a real dataset under classical and robust procedures. The *carbig* dataset () that contains various measured variables for about 392 automobiles. The $p = 5$ variables are the acceleration (X1), Displacement (X2), horsepower (X3), MPG (X4), and weight (X5). The summary of the factor loadings and variance explained under various procedures are listed in the table 1 and the factor loadings with 2% contamination are given in the table 2 which are given in the appendix. From the factor analysis, for the given data points there are two factors are extracted by all classical and robust procedures. It is observed from the table 1 the robust procedure also produces the same results as classical. For the contaminated data the deviation of factor loadings are very low in robust procedures but not in the case of classical procedures. The bi-plots of the factor loadings under various procedures with and without contamination displayed in the figure 1 and 2 respectively. It is observed that, all bi-plots based on the robust procedures with and without contamination is almost same, but in case of classical procedure the bi-plot shows the difference.

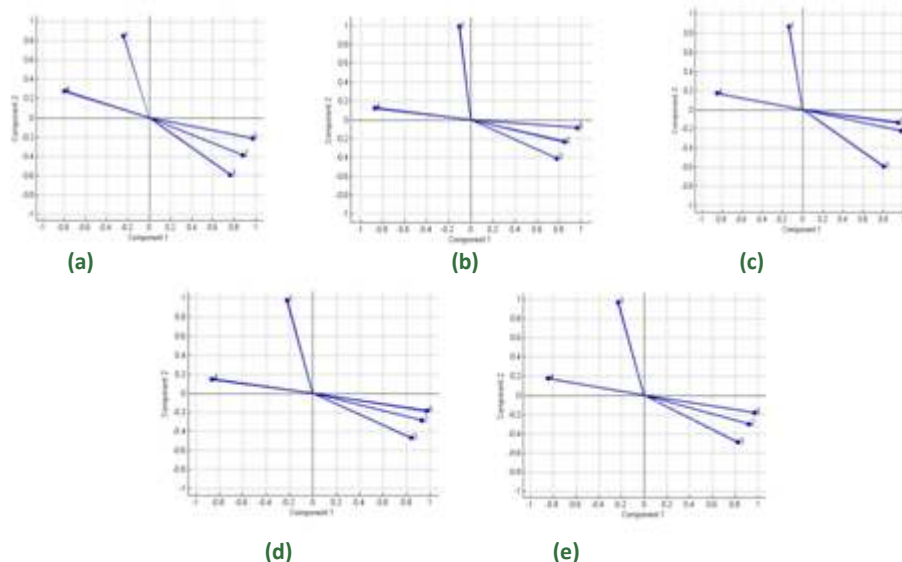


Figure 1: Bi-Plot

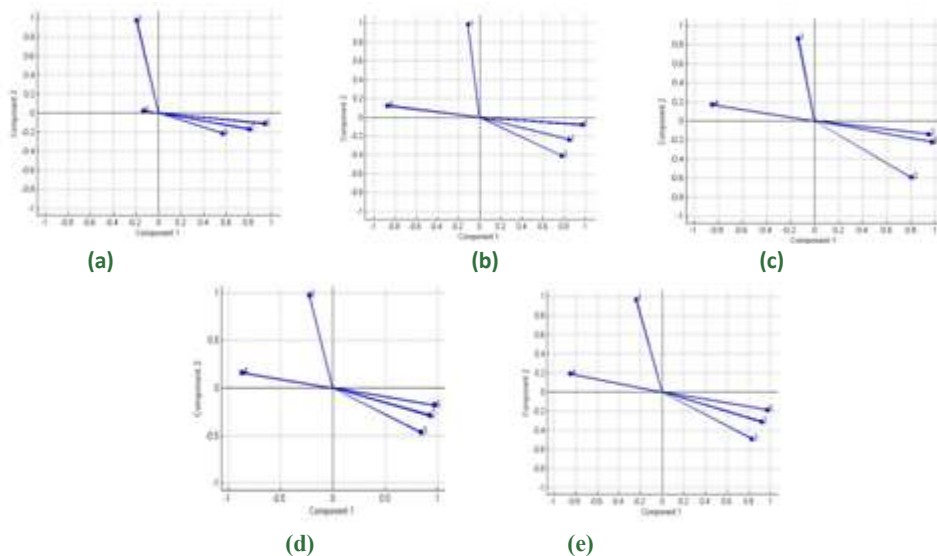


Figure 2: Bi-Plot (With Contamination) (a) Classical (b) MCD (c) MVE (d) S (e) MM

Experiment 2

The Olympic decathlon dataset is considered (see Linden 1987) for the experiment. The dataset description is as follows: the dataset contains the performances of 33 men's decathlon at the Olympic Games (1988) with ten different events. The ten different events are as follows 100 meters (Y1), long jump (Y2), shot-put (Y3), high jump (Y4), 400 meters (Y5), 110-meter hurdles (Y6), discus throw (Y7), pole vault (Y8), javelin (Y9) and 1500 meters (Y10). The factor analysis results for the given dataset and the results under various level of contamination (2%, 5%, 10% and 20%) of the data are displayed in the tables 3 to 7 which are given in the appendix. It is observed from the factor analysis results, for the given dataset there are three factors are extracted by the classical and robust procedures. Table 3 indicates that almost all the procedures classified the factor along with variables are same. The robust procedure gives the same results. Factor 1 contains 3 variables; they are 100 meters (Y1), 110 hurdles (Y6) and 400 meter (Y5). Factor 2 contains six variables like Long jump (Y2), Shot-put (Y3), High jump (Y4), Discuss throw (Y7), Pole vault (Y8) and Javelin throw (Y9). Factor 3 has only one variable, 1500 meters (Y10) running. Three factors can be named as sprints, field events and middle distance respectively. The results based on various levels of contamination of data are displayed in the tables 4 to 7. It is observed that the classical procedure doesn't extract the same variables along with factors. The contamination level was increased the classical procedure doesn't to classify the variables in a correct manner. The robust procedures, MCD and MVE are classified the variables in the factors in a meaningful way up to 35% of the contamination level, since these two procedures based on robust distance. But S and MM robust procedures tolerate up to some lower level of contamination of the data, because these two procedures are based on the magnitude of the Mahalanobis distance.

CONCLUSION

Robust location and scatter estimators find numerous applications to multivariate data analysis and inference in turn it play an important role in many areas such as pattern recognition, telecommunication applications, signal processing and computer vision tasks. In this context, this paper proposed to construct factor analysis with the help of most widely used robust estimators MVE, S and MM that can resist the effect of contaminated data. It is observed from the proposed factor analysis results, the classical procedure and robust procedures extract the same variables along with factors. The contamination level was increased the classical procedure doesn't classify the variables in the correct manner with a factor. The robust procedures can tolerate some level of contaminated data.

ACKNOWLEDGEMENT

First author convey his sincere thanks to University Grants Commission, New Delhi, India for providing financial assistance under the major research project [F.N.40-247/2011 (SR)] scheme awarded at the department of statistics, Bharathiar University, Coimbatore - 641046, Tamilnadu, India.

REFERENCES

1. Campbell, N.A., "Robust Procedures in Multivariate Analysis I: Robust covariance estimation", Applied Statistics, 29, 231-237, 1980.
2. Davies, P.L. "Asymptotic Behavior of S-Estimates of Multivariate Location Parameters and Dispersion Matrices", Annals of Statistics, 15, 3, 1269-1292, 1987.
3. Flury, B. and Riedwyl, H., "Multivariate statistics: a practical approach", Cambridge university press, 1988.
4. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A., "Robust Statistics: The Approach Based in Influence Functions", John Wiley and Sons, New York, 1986.
5. Huber, P.J., "Robust Statistics", John Wiley and Sons, New York, 1981.
6. Kent, J.T. and Tyler, D.E., "Constrained M-Estimation for Multivariate Location and Scatter", Annals of Statistics, 24, 1346-1370, 1996.
7. Kosfeld, R., "Robust exploratory factor analysis", Statistical papers, 37, 105-122, 1996.
8. Lopuhaa, H.P., "On the Relation between S-Estimators and M-Estimators of Multivariate Location and Covariance", Annals of Statistics, 17, 1662-1683, 1989.
9. Lopuhaa, H.P. and Rousseeuw, P.J., "Breakdown Properties of Affine Equivariant Estimators of Multivariate Location and Covariance Matrices", Annals of Statistics, 19, 229-248, 1991.
10. Maronna, R.A., "Robust M-estimation of Multivariate Location and Scatter", The Annals of Statistics, 4, 51-67, 1976.
11. Rousseeuw, P.J., "Least Median of Squares Regression", Journal of the American Statistical Association, 79, 871-880, 1984.
12. Rousseeuw, P.J., "Multivariate Estimation with High Breakdown Point", Mathematical Statistics and Applications, 283-297, 1985.
13. Rousseeuw, P.J. and Leroy, A., "Robust Regression and outlier detection", John Wiley and Sons, New York, 1987.
14. Rousseeuw, P.J. and Van Zomeren, B. C., "Unmasking Multivariate Outliers and Leverage Points", Journal of the American Statistical Association, 85, 633 – 639, 1990a.
15. Rousseeuw, P.J. and Van Zomeren, B. C., "Unmasking Multivariate Outliers and Leverage Points (With Discussion)", Journal of the American Statistical Association, 85, 633-651, 1990b.
16. Rousseeuw, P.J. and Van Zomeren, B. C., "Robust Distance: Simulation and Cutoff Values", Directions in Robust Statistics and Diagnostics, Part II, eds, W. Stahel and S. Welsberg, The IMA volumes in Mathematics and its Application, 34, 195-203, 1991.
17. Rousseeuw, P. J. and Van Driessen, K., "A Fast Algorithm for the Minimum Covariance Determinant Estimator", Technometrics, 41, 212-223, 1999.
18. Rousseeuw, P.J. and Yohai, V.J., "Robust Regression by Means of S- Estimators", Robust and Nonlinear Time Series Analysis Lecture Notes in Statistics, 26, 256-272, 1984.
19. Salibián-Barrera, M. and Yohai, V. J., "A fast algorithm for S-regression estimates", Journal of Computational and Graphical Statistics, 15, 414–427, 2006.
20. Tyler, D.E., "A class of asymptotic tests for principal component vectors", Annals of Statistics, 11(4), 1243-1250, 1983.
21. Tyler, D.E., "Some results on the existence and computation of the M-estimates of multivariate location and scatter", SIAM J. Sci. Stat. Comput., 9, 2, 354-362, 1988.
22. Tyler, D.E., "Some issues in the robust estimation of multivariate location and scatter", in Directions in Robust Statistics and Diagnostics Part II, Stahel, W. and Weisberg, S. (eds.), The IMA Volumes in Mathematics and its Applications, Springer-Verlag: New York, 34, 327-336, 1991.

Appendix
Table 1: Factor Loadings

Variables	Classical		MCD		MVE		S		MM	
X1	-0.2432	-0.8500	-0.1042	0.9920	-0.1365	0.8653	-0.2193	0.9731	-0.2298	0.9707
X2	0.8773	0.3871	0.8469	-0.2348	0.9434	-0.1374	0.9301	-0.2825	0.9213	-0.3005
X3	0.7618	0.5930	0.7758	-0.4101	0.8019	-0.5933	0.8424	-0.4682	0.8266	-0.4922
X4	-0.7978	-0.2786	-0.8705	0.1262	-0.8491	0.1706	-0.8678	0.1489	-0.8487	0.1777
X5	0.9692	0.2129	0.9635	-0.0847	0.9728	-0.2210	0.9724	-0.1864	0.9698	-0.1829
Variance Explained	99.7554	99.9616	99.7670	99.9084	99.9165	99.9835	99.8652	99.9769	99.8300	99.9727

Table 2: Factor Loadings (with 2% contamination)

Variables	Classical		MCD		MVE		S		MM	
X1	-0.1915	0.9789	-0.1123	0.9875	-0.1363	0.8650	-0.2247	0.9719	-0.2394	0.9684
X2	0.8014	-0.1691	0.8445	-0.2352	0.9423	-0.1376	0.9284	-0.2887	0.9190	-0.3103
X3	0.5682	-0.2115	0.7763	-0.4098	0.8013	-0.5929	0.8448	-0.4649	0.8284	-0.4893
X4	-0.1316	0.0236	-0.8725	0.1260	-0.8489	0.1700	-0.8692	0.1607	-0.8497	0.1917
X5	0.9399	-0.1128	0.9693	-0.0823	0.9725	-0.2213	0.9724	-0.1815	0.9690	-0.1840
Variance Explained	98.6157	99.3428	99.8607	99.9719	99.9006	99.9908	99.8593	99.9773	99.8263	99.9734

Table 3: Factor Loadings (Olympic Decathlon Data)

Events	Factor Analysis (FA)			MCD based FA			MVE based FA			S based FA			MM based FA		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
100 meters	0.7838	-0.0559	0.0708	0.7758	-0.1569	0.0462	0.7821	-0.0855	0.0624	0.7758	-0.157	0.0462	0.7821	-0.0855	0.0624
Long jump	-0.6091	0.0502	-0.2622	-0.5035	0.2072	-0.1241	-0.6132	0.089	-0.2022	-0.5035	0.2072	-0.1241	-0.6132	0.089	-0.2022
Shot-put	-0.2062	0.9687	0.1189	-0.1557	0.9732	0.1541	-0.1808	0.9684	0.1566	-0.1557	0.9732	0.1541	-0.1808	0.9684	0.1566
High jump	-0.2525	0.0827	-0.0691	-0.288	0.0204	-0.0313	-0.2629	0.0217	-0.0341	-0.2881	0.0204	-0.0313	-0.2629	0.0217	-0.0341
400 meters	0.7236	0.205	0.3746	0.7047	0.085	0.3106	0.7156	0.1922	0.3559	0.7047	0.085	0.3106	0.7156	0.1922	0.3559
110 hurdles	0.826	-0.1223	-0.0515	0.7286	-0.3996	-0.1412	0.8069	-0.1939	-0.0462	0.7286	-0.3996	-0.1412	0.8069	-0.1939	-0.0462
Discusthrow	-0.0674	0.7852	0.2645	-0.1944	0.734	0.3245	-0.0928	0.7492	0.34	-0.1944	0.734	0.3245	-0.0928	0.7492	0.34
Pole vault	-0.5437	0.376	0.0319	-0.5566	0.4249	0.012	-0.5645	0.3869	0.0003	-0.5566	0.425	0.012	-0.5645	0.3869	0.0003
Javelin	-0.0305	0.6143	-0.0324	-0.0901	0.5883	-0.1457	-0.0311	0.6273	-0.0775	-0.0901	0.5883	-0.1457	-0.0311	0.6273	-0.0775
1500 meter	0.2644	0.2189	0.9366	0.2197	0.0977	0.9681	0.2613	0.1712	0.9473	0.2197	0.0977	0.9681	0.2613	0.1712	0.9473
Variance	81.1567	95.5540	99.2999	84.4668	96.3167	99.4672	81.8294	93.2335	99.3417	81.5071	96.2013	99.3272	81.3661	96.0465	99.3277

Table 4: Factor Loadings (Olympic Decathlon with 2% contamination)

Events	Factor Analysis (FA)			MCD based FA			MVE based FA			S based FA			MM based FA		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
100 meters	0.8205	0.3949	-0.4014	-0.2414	0.4735	-0.3606	-0.0595	0.765	-0.3529	0.7841	-0.1998	0.0159	0.7815	-0.1475	0.0391
Long jump	0.5049	0.7401	-0.1745	0.2186	-0.5142	-0.1136	0.3154	-0.164	0.8423	-0.5138	0.2044	-0.1532	-0.6111	0.1021	-0.2169
Shot-put	-0.0226	-0.1777	0.9813	0.8414	-0.0454	0.0635	0.9836	0.0689	0.151	-0.1103	0.9829	0.1291	-0.135	0.9746	0.1641
High jump	0.7844	0.5068	-0.3511	-0.0118	0.0053	0.9974	-0.0414	-0.4912	0.0982	-0.3905	-0.0428	-0.0718	-0.3562	-0.0716	-0.0718
400 meters	-0.5246	-0.7583	0.3015	-0.0905	0.7889	-0.0167	0.2292	0.4972	-0.4495	0.701	0.0346	0.2976	0.7227	0.1188	0.345
110 hurdles	-0.7314	-0.6086	0.2744	-0.6019	0.4717	-0.3033	-0.4414	0.8872	0.1142	0.7165	-0.4397	-0.161	0.7994	-0.2643	-0.0698
Discusthrow	0.7885	0.3994	0.0079	0.9777	-0.0391	-0.1277	0.8114	-0.1323	0.0499	-0.1399	0.7396	0.306	-0.0546	0.7337	0.3559
Pole vault	0.64	0.6301	0.006	0.697	-0.6181	0.2597	0.4841	0.0191	0.5702	-0.5296	0.4562	0.016	-0.5406	0.4273	0.0136
Javelin	0.7517	0.3797	0.0729	0.3714	-0.279	0.0313	0.3568	-0.078	0.4333	-0.0699	0.5745	-0.2109	-0.0136	0.6099	-0.1348
1500 meter	-0.3972	-0.6727	0.3737	0.4136	0.5829	-0.2243	0.0911	0.2005	-0.5277	0.2513	0.0873	0.9614	0.2803	0.121	0.9496
Variance	86.1194	97.1504	99.4150	83.4165	96.1903	99.4219	77.1010	93.8564	99.2590	82.6629	95.6706	99.2971	82.6728	96.0101	99.2980

Table 5: Factor Loadings (Olympic Decathlon with 5% contamination)

Events	Factor Analysis (FA)			MCD based FA			MVE based FA			S based FA			MM based FA		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
100 meters	0.8691	0.4904	-0.0368	-0.1134	0.6706	-0.446	0.7009	0.0745	-0.1521	0.714	-0.1453	0.1199	0.746	-0.1148	0.0475
Long jump	0.034	-0.9376	-0.0061	0.0085	-0.0837	0.9939	-0.9683	0.2066	0.121	-0.5537	0.0722	-0.2461	-0.6033	0.0232	-0.1823
Shot-put	0.2043	0.6938	0.687	0.7946	-0.207	0.1527	0.0792	0.9217	0.2438	-0.16	0.9496	0.2603	-0.1346	0.9674	0.2028
High jump	0.9909	0.1102	-0.0458	-0.1698	-0.5626	-0.2637	-0.1573	-0.0794	0.3571	-0.4296	-0.1194	-0.022	-0.3625	-0.0728	-0.0572
400 meters	-0.9393	-0.2168	0.058	-0.1127	0.5899	-0.238	0.772	0.3198	0.112	0.6378	0.091	0.4087	0.6843	0.1404	0.3583
110 hurdles	0.3713	0.926	-0.0059	-0.499	0.7301	-0.199	0.7299	-0.0608	-0.5644	0.896	-0.274	-0.07	0.8725	-0.2488	-0.0418
Discusthrow	0.9514	0.0341	0.2143	0.9811	0.0066	0.1799	0.3661	0.7957	0.002	-0.1005	0.6783	0.4514	-0.044	0.7214	0.3867
Pole vault	0.9208	0.2292	0.1209	0.5999	-0.5205	0.3723	0.0137	0.2088	0.7707	-0.5742	0.3746	0.0696	-0.5612	0.3732	0.0861
Javelin	0.7529	-0.2732	0.3037	0.2314	-0.1391	0.495	-0.1848	0.711	-0.0354	0.0041	0.6482	-0.1747	0.0104	0.6254	-0.1195
1500 meter	0.0679	0.9026	0.1618	0.4474	0.3793	-0.0955	0.5254	0.3973	0.5088	0.2044	0.0952	0.849	0.2432	0.1428	0.9568
Variance	80.4564	97.0798	98.8565	83.4165	96.1903	99.4219	84.7222	97.2415	99.3698	82.3514	96.0679	99.2817	82.1740	96.0542	99.2839

Table 6: Factor Loadings (Olympic Decathlon Data with 10% contamination)

Events	Factor Analysis (FA)			MCD based FA			MVE based FA			S based FA			MM based FA		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
100 meters	0.7361	-0.3822	0.5524	-0.368	0.5799	-0.277	-0.2318	0.5912	-0.1555	-0.2225	0.5179	-0.5	-0.1912	0.5074	0.5032
Long jump	0.7453	0.6567	-0.0964	0.3318	-0.3094	-0.1297	0.2108	-0.342	0.4974	0.0479	-0.3829	0.4195	0.0512	-0.5132	-0.376
Shot-put	0.5685	0.3777	0.5458	0.9156	0.0848	0.045	0.8488	0.1056	-0.0811	0.9814	0.0923	0.1529	0.9697	0.1317	-0.1931
High jump	0.7939	0.5988	0.0904	0.0358	-0.0801	0.9936	-0.361	-0.0137	0.9298	-0.1654	0.0762	0.5981	-0.087	-0.0244	-0.3984
400 meters	0.0213	0.9655	-0.1713	-0.0789	0.8983	0.0934	0.1624	0.9667	-0.1848	0.0746	0.9735	-0.2041	0.1087	0.8037	0.3187
110 hurdles	0.1814	-0.2353	0.9522	-0.7241	0.4288	-0.2334	-0.2806	0.4521	-0.7347	-0.3069	0.3708	-0.8195	-0.2071	0.319	0.9222
Discusthrow	0.9281	-0.219	0.135	0.8924	0.0512	-0.155	0.8601	0.1804	-0.1059	0.7578	0.1253	0.0599	0.7476	0.2438	-0.0943
Pole vault	0.2253	-0.9054	0.2403	0.7077	-0.491	0.2327	0.353	-0.2705	0.2487	0.3807	-0.2575	0.4866	0.3626	-0.2238	-0.4876
Javelin	0.8603	-0.0587	-0.1298	0.5753	-0.0169	0.0461	0.6793	-0.2356	0.2279	0.5972	0.0034	-0.0177	0.6408	-0.0837	0.0811
1500 meter	-0.1942	-0.168	0.8788	0.2406	0.5842	-0.1573	0.1494	0.7846	-0.1483	0.1976	0.5636	-0.0196	0.1798	0.6937	-0.0106
Variance	61.8370	90.3361	97.9152	82.3887	96.4301	99.3959	83.0026	94.8756	99.2213	81.4511	96.1185	99.2509	81.1121	95.9612	99.2627

Table 7: Factor Loadings (Olympic Decathlon Data with 20% contamination)

Events	Factor Analysis (FA)			MCD based FA			MVE based FA			S based FA			MM based FA		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
100 meters	0.2271	0.8423	-0.4837	-0.2234	0.9354	-0.2649	0.8059	0.3505	-0.0737	-0.2663	0.5838	0.4644	-0.2674	0.5714	0.4694
Long jump	0.9853	-0.1188	-0.1014	0.137	-0.1943	0.9688	-0.6051	0.1176	-0.6585	0.1217	-0.4522	-0.3132	0.1145	-0.4662	-0.3251
Shot-put	0.1059	0.5656	0.0908	0.9565	-0.009	0.0839	0.1025	0.9868	-0.1037	0.9926	0.0527	-0.083	0.9926	0.0557	-0.0818
High jump	0.9876	0.1088	0.0917	0.0787	-0.3484	-0.2228	-0.4041	0.3754	-0.1362	-0.1511	-0.0443	-0.6251	-0.1543	-0.0524	-0.6231
400 meters	0.2	-0.2014	0.9312	-0.0237	0.6634	-0.0491	0.6612	0.1519	0.3208	0.0114	0.8166	0.2644	0.0099	0.7997	0.283
110 hurdles	-0.0665	0.9208	0.1358	-0.7479	0.5535	-0.0401	0.7661	-0.3027	-0.0439	-0.4866	0.2557	0.8324	-0.4821	0.2486	0.8372
Discusthrow	0.5356	0.4535	-0.2864	0.7961	0.1123	0.0851	0.0544	0.7432	0.0315	0.7406	0.1834	-0.0854	0.7396	0.1996	-0.0792
Pole vault	-0.3136	0.5604	-0.008	0.6185	-0.204	0.1709	-0.1932	0.3917	0.3471	0.4819	-0.1615	-0.3935	0.4814	-0.1468	-0.3973
Javelin	0.5483	-0.0373	0.0398	0.4957	0.0084	0.563	-0.112	0.2834	-0.5158	0.5498	-0.1753	0.0652	0.5479	-0.1896	0.0607
1500 meter	-0.2032	0.4102	0.8863	0.1865	0.2656	-0.1161	0.0358	0.1903	0.967	0.0996	0.6348	-0.1138	0.1026	0.655	-0.109
Variance	83.7221	91.3759	96.6906	79.8832	95.2993	99.2132	66.8015	89.2205	98.8539	81.1217	95.4153	99.1854	81.1973	95.3950	99.1796

Source of Support: None Declared
Conflict of Interest: None Declared