JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

Data Depth based Discriminant Classification Analysis

[*1]Ramkumar N, [2]Nazia Wahid, [3]Yookesh.T.L and [4]Keerthika .K.S

[*1]Department of Community Medicine, Dayananda Sagar University, Bangalore, India

[2]Department of Mathematics and Statistics, Vishwakarma University, Pune, Maharashtra,India

[3,4]Department of Mathematics, Vignan's Foundation for Science, Technology and Research, Guntur, Andhra Pradesh, India

E.Mail: [*1]ram222u@gmail.com, [2]wahidnazia010@gmail.com, [3]renu_yookesh@yahoo.co.in and [4]keerthikasundaram.29@gmail.com

Received 2022 April 02**; Revised** 2022 May 20**; Accepted** 2022 June 18**.**

Abstract:

The Data Depth is used to measure the depth or area of any variation according to the distribution base. This results in the average natural centre-outer of the sample points. The essence of the deep procedure in multivariate analysis is to measure the degree of centrality of points associated with assumptions or probability distributions. This working data examines in-depth methods for determining the size of the site, ie. deepest or focal point. In addition, various in-depth procedures are studied in real and simulation contexts using R software. The performance of various data-depth processes is analyzed with numerical description by calculating the average misclassification error as part of a discriminative analysis.

Keywords: Data Depth, Location, Scatter and Linear discriminant analysis.

1.1 Introduction

The discipline statistics contributes almost all the fields, either directly or indirectly. In statistics, measure of location is extremely important for univariate / multivariate data analysis techniques. The conventional sample mean (vector) is very sensitive when the data contains extremes and thus gives the unreliable estimate of the population mean. For the past few decades, a substantial growth in statistics, specifically, in the context of estimation of measure of location such as robust based statistics, depth-based statistics etc. Now-a-days, the concept of depth in statistics attracts the researchers, because it gives the reliable estimates of location in a given data cloud. This chapter provides some preliminaries on data depth, development of data depth and also presents an overview of this dissertation.

1.2 Data Depth

Depth is an integer value that matches the specified candidate record. This results in an outside-inside/center-outside array sampling points. Normal order items are changed from the highest order. In a typical statistical table, the data is organized from the smallest to the largest sample point, but the statistical depth starts in the middle of the sample and extends in all directions.

Data depth is a major concept from nonparametric tends to multivariate data analysis. There is one possible way of ordering the multivariate data, specifically to a centraloutward ordering. Data depth is basically a position of the data point in whole data points in data cloud. The depth of a point is relative to the 'deepest' point in a given data cloud. The data depth isprovides center-outward ordering of points in any dimension and leads to a new non-parametric multivariate statistical analysis in which no distributional assumption is needed.Nonparametric analysis relies heavily on signs and ranks, order statistics, quantiles, and outlyingness functions.
In principle, any function that provides a reasonable center-outside ordering of points in multidimensional space can be considered a depth function. Based on depth functions, methods of signs and ranks, order statistics, quantiles, and distance measures could be conveniently extended from a multivariate framework in a unified way. These functions form a basis for the detection of eccentricity contours, taking into account the geometry of the data.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

1.3 Depth Contour

The depth line is a line on a nautical chart that connects points of equal depth. A contour of a function of two variables is a curve along which the function has a constant value, such that the curve connects points of equal value. It is a planar section of the three-dimensional graph of the function f(x,y) parallel to the x,y-plane. Contour lines are curved, straight, or a mixture of the two lines on a map that describe the intersection of a real or hypothetical surface with one or more horizontal planes. The configuration of these contours allows map readers to derive the relative gradient of a parameter and to estimate that parameter at specific locations.

2.Data Depth Procedures

A variety of graphic and quantitative methods are defined for indices such as location, size, and shape, as well as to compare inference methods based on data depth. In recent decades, many concepts of depth have been proposed. The known depth methods such as Mahalanobis depth [1], Half Space Depth [2], Simplicial Depth [4], Simplicial Block Depth [3], Spatial Depth [12], Zonoid Depth [8], Projection Depth [13, 15] are summarized in this section.

2.1 Half Space Depth

It is introduced by Tukey in 1975. The depth of the point Half space $x = (x_1,....,x_p) \in S_n = \{x_i = (x_{i1},.....,x_{ip}); i = 1,...,n\} \subset \mathbb{R}^p$ with respect to relative to a p - dimensional data set S_n is defined as the minimum number of data points in a closed half-space bounded by x. In the one-dimensional case, it is easy to see that the depth of a point is determined by the expression, $\min\{\#\{x_i \leq x\}, \#\{x_i \geq x\}\}$ the median is the point (or points) with maximal depth. In diversity, the median can be absolute because it has the greatest depth. This transition is called “Tukey median”. HSD is also known as “Tukey depth and Local depth”.

2.2 Mahalanobis Depth

The concept, generalized distance in statistics is given by Mahalanobis (1936). In 1975, the Mahalanobis distance was used as a measure to calculate the depth of a point. MD of a point $x \in S_n \subset \mathbb{R}^p$ relative to a p -dimensional data set defined as:

$$MD(x; S_n) = \left[1 + \left(x - \bar{x}\right)^T S^{-1}\left(x - \bar{x}\right)\right]^{-1} \qquad (1)$$

where $\bar{x}$ and S are the mean vector and dispersion matrix of S_n.

This function is unreliable because it relies on unreliable measures such as mean and variance matrix. Another disadvantage of this procedure is that it depends on the continuity of the second instants.

2.3 Projection Depth

Let μ(.) and σ(.) be univariate location and scale events, respectively. Then the outlyingness of a point with deference to the distribution function *F* of *x* defined by (Liu 1992)

$$O(x, F) = \sup_{\|u\|=1} \left|Q(u, x, F)\right| \qquad (2)$$

here, $Q(u, x, F) = \left(u^T x - \mu(F_u)\right)/\sigma(F_u)$ and F_u is the distribution of $u^T x$. Let, μ (.) and σ (.) be multivariate case used a point of a p-dimensional data set. The projection depth (PD) is defined by

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

$$PD(x,F)=\frac{1}{1+O(x,F)} \quad (3)$$

2.4 Simplicial Depth

Liu (1990) introduced the concept of SD. This is the point $x \in S_n \subset \mathbb{R}^p$ respect to the data set of p-dimension S_n, defined as the number of closed simplexes containing x and having $p+1$ vertices in S_n. In the bivariate case, the simplicial depth of a point x is the number of triangles that passes through the vertices at S_n and contain x. SD is calculated as the probability that a point lies in a simplex built on d + 1 data points.

$$D_S(x,F)=P_F\left(x\in S[X_1,\ldots,X_{d+1}]\right), x\in R^d \quad (4)$$

Simple depth is strong against extreme values. This is because if a set of sample points is represented by a maximum depth point, it is possible to arbitrarily deform up to a specified range of sample points without substantially changing the position of the representative point. It does not change when the connection level changes. However, single depth has no other desirable properties for measuring strong central stresses. With Centro symmetric distributions, there is not necessarily a clear point of maximum depth at the center of the distribution. Also, from the maximum depth point, the simple depth does not necessarily decrease smoothly.

2.5 Simplicial Volume Depth

Oja (1983) established a depth procedure using the SVD. A simplicial volume is an invariant of the homotopy of associated closed oriented manifolds introduced by Gromov (1983). Intuitively, simplicial volume phenomena are difficult to describe in terms of the simplicity (with real coefficients) of the manifold we are considering.

Let M be an associated closed oriented manifold of dimension n. Then the simplicial volume of M (also called the Gromov norm of M) is defined as, $\|M\| := \|[M]\|_1 = \inf\{|c|_1 | c \in C_n(M;\Re)$ is a fundamental cycle of $M\} \in \Re_{\geq 0}$, where, $[M] \in H_n(M;\mathbb{R})$ is the fundamental class of M with real coefficients. Oja depth of a point $x \in S_n \subset \mathbb{R}^p$ relative to a p-dimensional data set S_n is defined as the sum of the volume of every closed simplex having a vertex at x and the others in any p points of the S_n data set. In the bivariate case, the Oja depth of a point x relative to a bivariate data set S_n is the sum of the areas of all triangles whose vertices are x, x_i, x_k with x_i and x_k belonging to S_n.

2.6 Zonoid Depth

Koshevoy and Mosler (1996) introduced a notion of data depth, called Zonoid Data Depth (ZD). The zonoid data depth, depth $_\mu$(x), of a point $X \in \mathbb{R}^d$ is defined by,

$$\mathrm{depth}_\mu(x)=\begin{cases}\sup\{\alpha : \mathrm{x}\in \mathrm{D}_\alpha(\mu)\}, & \text{if } \mathrm{x}\in \mathrm{D}_\alpha(\mu) \text{ for some } \alpha,\\ 0, & \text{otherwise.}\end{cases} \quad (5)$$

The data depth of a point x is the maximal height α at which $\alpha x \in proj_\alpha \hat{Z}(\mu)$. Here,

$$D_\alpha(\mu)=\frac{1}{\alpha}\mathrm{proj}_\alpha\left(\hat{Z}(\mu)\right) \quad (6)$$

where $0<\alpha\leq 1$. Further, the depth of x equals zero if x lies outside $D_\alpha(\mu)$ for all α; it equals one if x is the expectation. If α > 0, $D_\alpha(\mu)$ is the set of all points that include data depth greater than or equal to α.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

2.7 Spatial Depth

An implementation of the idea of spatial depth (SPD), established by Serfling (2002), which is defined as follows: Lt Y be d-dimensional random vectors have cumulative distribution function F. Then, the multivariate spatial depth of $x \in \mathbb{R}^d$ qualified F is defined as,

$$SD(x,F) = 1 - \left\| \int S(x-y)\,dF(y) \right\|_E = 1 - \left\| E\left[(x-y)\right] \right\|_E \qquad (7)$$

where $\|.\|_E$ is the Euclidean norm in $\mathbb{R}^d$. The spatial depth is a depth function that builds ahead the notion of spatial (also called geometric) quantiles for multivariate data, considered by Chaudhuri (1996) and Koltchinskii (1997), formulated by Vardiand Zhang(2000) and Serfling (2002). This Spatial depth also called L1-depth.

3.1 Computational Results

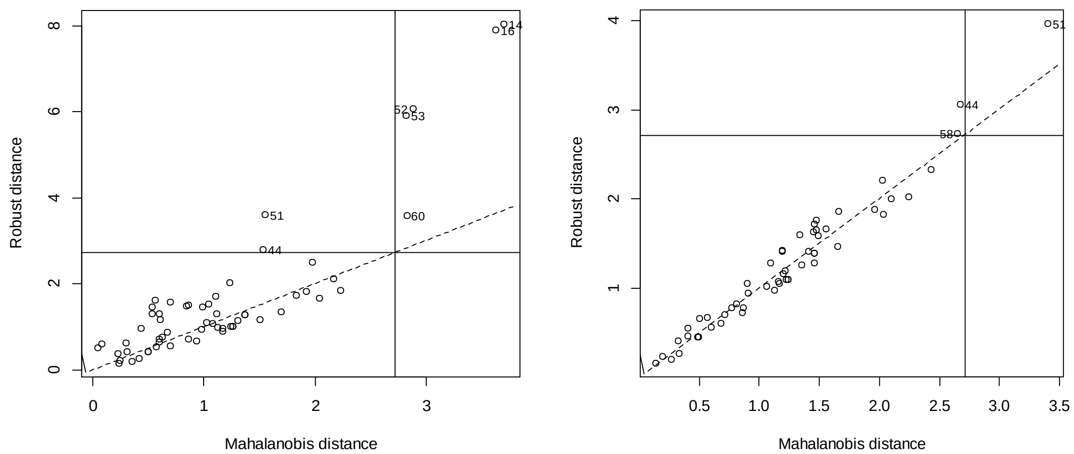
This session presents the performances of "Data Depth" procedures such as "Mahalanobis depth, Halfspace depth, Simplicial depth, Simplicial volume depth, Spatial depth, Zonoid depth and Projection depth" which are studied under real data and simulation. The results obtained from the study are summarized in the section 4.2 and 4.3 respectively. Further, the efficiency of data depth procedures has been studied by applying it into multivariate technique, specifically in the context of classification problems under real datasets and the results are summarized in the section 4.4.

3.2 Results on Real Data

This section provides the performance of various data depth procedures by presenting the results of numerical representations performed under the actual data set and accounting for them with/without outliers.

Case 1

Data Description: For this study, a real data set was considered, namely cardata90, subset from data on cars (Chambers and Hastie (1993)) (Appendix: A1). The data set contains two variables, with 60 observations. The variables are weight and engine displacement of cars. For the given data set, the 14th, 16th, 44th, 51st, 52nd, 53rd, 58th and 60th observations are identified as outliers through distance-distance plot (figure 3.1). The computed depth values and depth contour plots for all the observations (with/without outliers) under various depth procedures.The deepest point is located under various notions of depth procedures with and without outliers and is summarized in the table 3.1.



JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

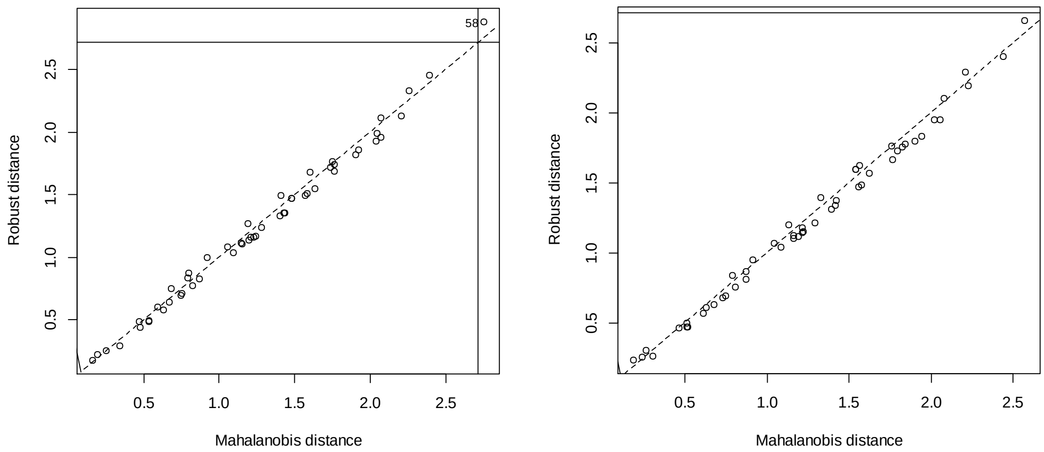


Figure 3.1: Distance-Distance Plots (with/without outlier) (cardata90)

Table 3.1: Measure of location and the associated depth value under various data depth procedures

Methods	MD	HSD	SD	SVD	SPD	ZD	PD
With Outlier	39 (2880,151) **0.998**	29 (2780,133) **0.350**	39 (2880,151) **0.282**	15 (2285,153) **0.766**	45 (2885,143) **0.859**	15 (2285,153) **0.963**	29 (2780,133) **0.648**
Without Outlier	45 (2885,143) **0.966**	29 (2780,133) **0.385**	29 (2780,133) **0.293**	40 (2975,153) **0.681**	29 (2780,133) **0.868**	45 (2885,143) **0.907**	29 (2780,133) **0.663**

. – Observation number**; (.) –** Location**; Bold –** Depth value

In the table above, when we consider the value of the maximum depth, we notice that the half-space depth (HSD) and the projection (PD) provide the same depth point (position measurement) with and without values aberrant. Both of these methods work better than the other methods. After removing the outliers, “simplicial (ST) and simplicial (SVT) volume depths” yield the same location as the HST and PD. “Zonoid (ZD) and Mahalanobis (MD)” depth do not provide reliable location measurements (deep point).

Case 2

Data Description: For this study, a set of real data was considered, namely data on delivery times (Montgomery and Beck (1982), p.116). This dataset consists of three variables with 25 observations. The variables are the number of products (x1), the distance (x2) and the delivery time (x3). For a given data set, the 9th, 11th, 20th and 22nd observations are identified externally by distance plots (Figure 3.2). Calculated depth values for all observations with and without outliers. The deep point lies under various concepts of deep procedures with and without outliers and is summarized in Table 3.2

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

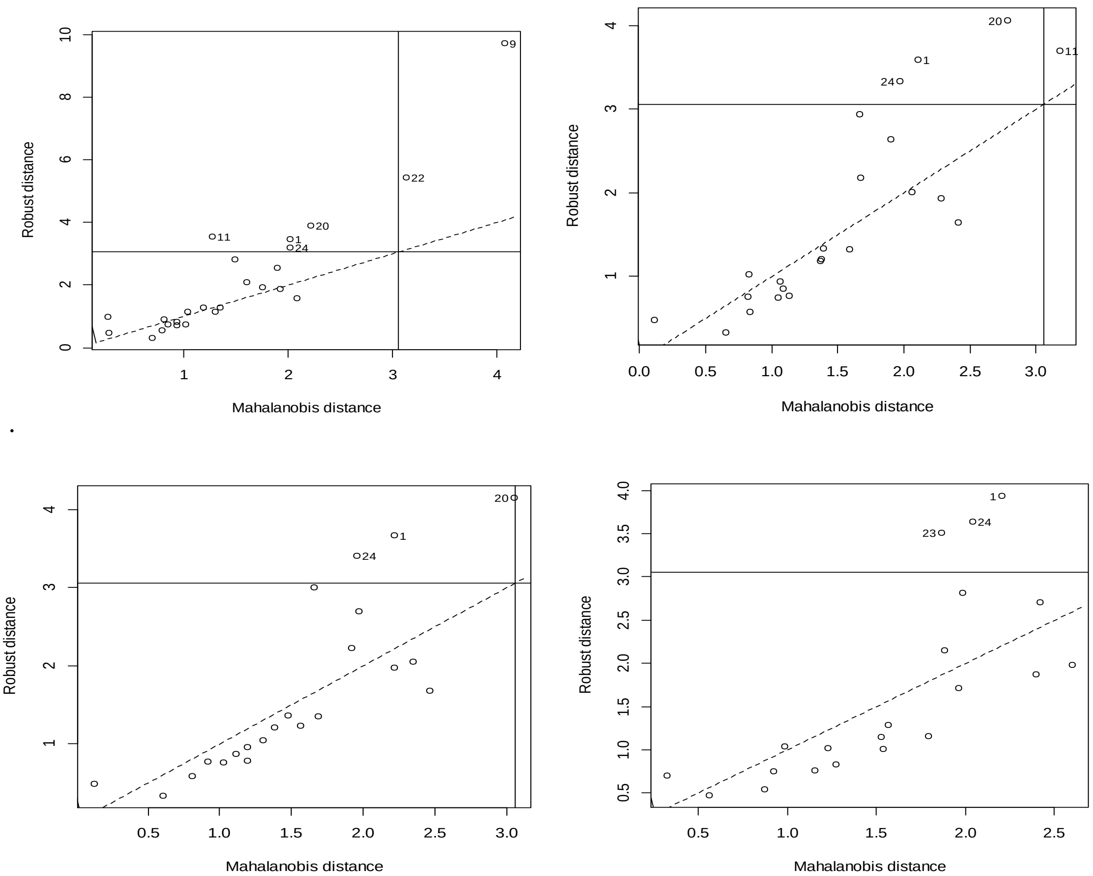


Figure 3.2: Distance-Distance Plots (with/without outliers) (delivery time data)

Table 3.2: Measure of location and the associated depth value under various data depth procedures

Methods	MD	HSD	SD	SVD	SPD	ZD	PD
With Outlier	15 (9,448,24) **0.932**	6 (7,330,18.11) **0.4**	15 (9,448,24) **0.252**	7 (2,110,8) **0.756**	6 (7,330,18.11) **0.859**	15 (9,448,24) **0.771**	6 (7,330,18.11) **0.605**
Without Outlier	6 (7,330,18.11) **0.905**	6 (7,330,18.11) **0.333**	19 (3,36,9.5) **0.303**	25 (4,150,10.75) **0.783**	6 (7,330,18.11) **0.772**	17 (6,200,15.35) **0.683**	6 (7,330,18.11) **0.5**

. – Observation number**; (.) –** Location**; Bold –** Depth value

From the above table, it is noticed that halfspace, spatial, projection depth performed well by comparing all the other depth procedures, since it gives the same location under with and without outliers. It is concluded that these procedures are robust in nature.

3.3 Results on Simulation

This section presents the performance of various data mining procedures by presenting results performed on simulated data with different levels of contamination. Also, various levels of pollution are considered with three categories namely location, quantity and location and scale of pollution.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

3.3.1 Location Contamination

Case 1

This section presents the results of a simulation study with location contaminations. For this study, the data were simulated from (n=50) normal distribution, mean vector μ=(0,0), and unit covariance matrix, $\Sigma=I_2$. The various level of contaminations (mean vector, μ=(4,4) and unit covariance matrix, $\Sigma=I_2$) such as 0%, 1%, 2%, 5%, 10%, 20% and 25% are considered and the obtained results are summarized in the table 3.3.

Table 3.3: Measure of location and the associated depth value under various data depth procedures

Error	MD	HSD	SD	SVD	SPD	ZD	PRD
0%	21 (0.220, -0438) **0.885**	21 (0.220, -0438) **0.34**	21 (0.220, -0438) **0.278**	21 (0.220, -0438) **0.648**	13 (0.236, 0.345) **0.776**	21 (0.220, -0438) **0.760**	10 (0.527, 0.016) **0.632**
1%	21 (0.220, -0438) **0.881**	21 (0.220, -0438) **0.34**	10 (0.527, 0.016) **0.282**	13 (0.236, 0.345) **0.687**	10 (0.527, 0.016) **0.793**	21 (0.220, -0438) **0.760**	10 (0.527, 0.016) **0.643**
2%	13 (0.236, 0.345) **0.909**	13 (0.236, 0.345) **0.36**	13 (0.236, 0.345) **0.282**	35 (0.417, 0.365) **0.757**	13 (0.236, 0.345) **0.818**	13 (0.236, 0.345) **0.811**	13 (0.236, 0.345) **0.675**
5%	45 (-0.204, -0.406) **0.943**	13 (0.236, 0.345) **0.38**	10 (0.527, 0.016) **0.289**	7 (-0.026, 0.515) **0.766**	13 (0.236, 0.345) **0.832**	13 (0.236, 0.345) **0.891**	10 (0.527, 0.016) **0.683**
10%	35 (0.417, 0.365) **0.991**	15 (-1.364, 0.873) **0.4**	15 (-1.364, 0.873) **0.304**	18 (0.438, 1.497) **0.721**	15 (-1.364, 0.873) **0.971**	35 (0.417, 0.365) **0.931**	15 (-1.364, 0.873) **0.755**
15%	42 (0.726, 0.694) **0.996**	13 (0.236, 0.345) **0.36**	13 (0.236, 0.345) **0.298**	11 (0.205, 1.016) **0.711**	32 (0.413, 0.485) **0.899**	42 (0.726, 0.694) **0.975**	32 (0.413, 0.485) **0.729**
20%	49 (1.081, 1.159) **0.950**	31 (0.662, 0.232) **0.36**	31 (0.662, 0.232) **0.296**	29 (0.302, -0.726) **0.720**	31 (0.662, 0.232) **0.915**	31 (0.662, 0.232) **0.858**	31 (0.662, 0.232) **0.669**
25%	42 (0.726, 0.694) **0.967**	35 (0.417, 0.365) **0.38**	35 (0.417, 0.365) **0.300**	32 (0.413, 0.485) **0.767**	35 (0.417, 0.365) **0.898**	42 (0.726, 0.694) **0.916**	35 (0.417, 0.365) **0.718**

. – Observation number**; (.) –** Location**; Bold –** Depth value

"Mahalanobis, Sonoid, and Half-Space depths" tolerate a certain amount of contamination and yield the same depth point (position measurement). Although data contamination is reduced, other depth mechanisms do not tolerate and provide the same depth point.

Case 2:

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

This section presents the results of a simulation study. For this study, the data were simulated (n=100) from normal distribution with mean vector μ=(0, 0), and unit covariance matrix, $\Sigma=I_2$. The various level of contaminations (mean vector, μ=(4,4) and unit covariance matrix, $\Sigma=I_2$) such as 0%, 1%, 2%, 5%, 10%, 20% and 25% are considered and the obtained results are summarized given below.

Table 3.4: Measure of location and the associated depth value under various data depth procedures

Error	MD	HSD	SD	SVD	SPD	ZD	PRD
0%	41 (-0.178, 0.169) **0.968**	41 (-0.178, 0.169) **0.42**	41 (-0.178, 0.169) **0.275**	40 (0.258, 0.317) **0.688**	41 (-0.178, 0.169) **0.918**	41 (-0.178, 0.169) **0.901**	41 (-0.178, 0.169) **0.784**
1%	90 (0.235, 0.033) **0.966**	41 (-0.178, 0.169) **0.43**	41 (-0.178, 0.169) **0.275**	82 (0.414, 0.183) **0.706**	41 (-0.178, 0.169) **0.919**	41 (-0.178, 0.169) **0.906**	41 (-0.178, 0.169) **0.782**
2%	90 (0.235, 0.033) **0.966**	41 (-0.178, 0.169) **0.43**	41 (-0.178, 0.169) **0.275**	40 (0.258, 0.317) **0.696**	41 (-0.178, 0.169) **0.920**	41 (-0.178, 0.169) **0.892**	41 (-0.178, 0.169) **0.792**
5%	90 (0.235, 0.033) **0.981**	41 (-0.178, 0.169) **0.43**	41 (-0.178, 0.169) **0.275**	20 (0.294, 0.834) **0.711**	41 (-0.178, 0.169) **0.921**	90 (0.235, 0.033) **0.934**	41 (-0.178, 0.169) **0.762**
10%	3 (0.429, 0.506) **0.990**	90 (0.235, 0.033) **0.44**	90 (0.235, 0.033) **0.276**	75 (0.506, 0.347) **0.716**	90 (0.235, 0.033) **0.925**	75 (0.506, 0.347) **0.934**	90 (0.235, 0.033) **0.734**
15%	73 (0.514, 0.399) **0.993**	90 (0.235, 0.033) **0.42**	90 (0.235, 0.033) **0.277**	73 (0.514, 0.399) **0.685**	90 (0.235, 0.033) **0.928**	73 (0.514, 0.399) **0.964**	90 (0.235, 0.033) **0.748**
20%	73 (0.514, 0.399) **0.968**	55 (0.355, 0.052) **0.43**	55 (0.355, 0.052) **0.278**	3 (0.429, 0.506) **0.721**	55 (0.355, 0.052) **0.951**	73 (0.514, 0.399) **0.922**	90 (0.235, 0.033) **0.846**
25%	43 (0.850, 0.698) **0.977**	73 (0.514, 0.399) **0.43**	73 (0.514, 0.399) **0.276**	90 (0.235, 0.033) **0.696**	73 (0.514, 0.399) **0.959**	43 (0.850, 0.698) **0.938**	73 (0.514, 0.399) **0.710**

. – Observation number; **(.) –** Location; **Bold –** Depth value

"Half Depths, Simplicial, Spatial and Projection depths" tolerate a certain amount of contamination and give the same depth point (measure of location). Although data contamination is minimal, other depth procedures cannot tolerate and do not provide the same depth point.

3.3.2. Scale Contamination

Case 3:

This section presents the results of a simulation study. For this study, the simulated data (n=50) from normal distribution, mean vector μ= (0, 0), and unit covariance matrix, $\Sigma=I_2$. The various level of contaminations (mean vector, μ=(0,0) and unit covariance matrix, $\Sigma=1.5I_2$) such as 0%, 1%, 2%, 5%, 10%, 20% and 25% are considered and the obtained results are summarized as follows.

Table 3.5: Measure of location and the associated depth value under various data depth procedures

Error	MD	HSD	SD	SVD	SPD	ZD	PRD
0%	15 (-0.000, -0.344) **0.996**	15 (-0.000, -0.344) **0.42**	15 (-0.000, -0.344) **0.306**	15 (-0.000, -0.344) **0.691**	15 (-0.000, -0.344) **0.964**	15 (-0.000, -0.344) **0.964**	15 (-0.000, -0.344) **0.731**
1%	15 (-0.000, -0.344) **0.995**	15 (-0.000, -0.344) **0.42**	15 (-0.000, -0.344) **0.306**	10 (-0.243, -0.486) **0.731**	15 (-0.000, -0.344) **0.951**	15 (-0.000, -0.344) **0.958**	15 (-0.000, -0.344) **0.745**
2%	15 (-0.000, -0.344) **0.981**	15 (-0.000, -0.344) **0.4**	15 (-0.000, -0.344) **0.305**	15 (-0.000, -0.344) **0.689**	15 (-0.000, -0.344) **0.952**	15 (-0.000, -0.344) **0.929**	15 (-0.000, -0.344) **0.713**
5%	35 (0.292, 0.220) **0.963**	15 (-0.000, -0.344) **0.42**	15 (-0.000, -0.344) **0.306**	24 (0.033, -0.650) **0.787**	15 (-0.000, -0.344) **0.965**	15 (-0.000, -0.344) **0.937**	15 (-0.000, -0.344) **0.745**
10%	31 (0.203, -0.268) **0.993**	31 (0.203, -0.268) **0.4**	15 (-0.000, -0.344) **0.305**	18 (0.698, -0.254) **0.702**	15 (-0.000, -0.344) **0.982**	31 (0.203, -0.268) **0.962**	31 (0.203, -0.268) **0.767**
15%	15 (-0.000, -0.344) **0.948**	15 (-0.000, -0.344) **0.34**	15 (-0.000, -0.344) **0.294**	41 (0.308, -0.724) **0.707**	15 (-0.000, -0.344) **0.844**	15 (-0.000, -0.344) **0.847**	15 (-0.000, -0.344) **0.636**
20%	29 (-0.078, -0.1250 **0.987**	28 (-0.203, -0.284) **0.38**	28 (-0.203, -0.284) **0.291**	10 (-0.243, -0.486) **0.713**	29 (-0.078, -0.1250 **0.873**	29 (-0.078, -0.1250 **0.938**	28 (-0.203, -0.284) **0.665**
25%	15 (-0.000, -0.344) **0.937**	15 (-0.000, -0.344) **0.42**	15 (-0.000, -0.344) **0.304**	32 (-.0.349, -0.375) **0.684**	15 (-0.000, -0.344) **0.932**	15 (-0.000, -0.344) **0.859**	15 (-0.000, -0.344) **0.749**

. – Observation number**; (.) –** Location**; Bold –** Depth value

“Simplicial and spatial depths” allow contamination up to 15% and are similar to point depth (measure of location). Other systems do not support depth, and although data contamination is more severe, they do not provide the same depth.

Case 4:

In this section results is based on simulation study. For this study, the data were simulated (n=100) from normal distribution, mean vector μ= (0, 0), and unit covariance matrix, $\Sigma=I_2$. The various level of contaminations (mean vector, μ=(0,0) and unit covariance matrix, $\Sigma=1.5I_2$) such as 0%, 1%, 2%, 5%, 10%, 20% and 25%, are considered and the obtained results are summarized in the table 3.6.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

Table 3.6: Measure of location and the associated depth value under various data depth procedures

Error	MD	HSD	SD	SVD	SPD	ZD	PRD
0%	67 (-0.191, -0.219) **0.967**	67 (-0.191, -0.219) **0.44**	67 (-0.191, -0.219) **0.277**	75 (-0.215, 0.325) **0.671**	67 (-0.191, -0.219) **0.947**	67 (-0.191, -0.219) **0.899**	67 (-0.191, -0.219) **0.759**
1%	67 (-0.191, -0.219) **0.968**	67 (-0.191, -0.219) **0.44**	67 (-0.191, -0.219) **0.278**	26 (-0.017, -0.418) **0.678**	67 (-0.191, -0.219) **0.955**	67 (-0.191, -0.219) **0.905**	67 (-0.191, -0.219) **0.767**
2%	67 (-0.191, -0.219) **0.956**	67 (-0.191, -0.219) **0.42**	67 (-0.191, -0.219) **0.275**	67 (-0.191, -0.219) **0.676**	67 (-0.191, -0.219) **0.926**	67 (-0.191, -0.219) **0.879**	67 (-0.191, -0.219) **0.713**
5%	67 (-0.191, -0.219) **0.966**	67 (-0.191, -0.219) **0.43**	67 (-0.191, -0.219) **0.277**	67 (-0.191, -0.219) **0.682**	67 (-0.191, -0.219) **0.946**	67 (-0.191, -0.219) **0.906**	67 (-0.191, -0.219) **0.780**
10%	67 (-0.191, -0.219) **0.958**	67 (-0.191, -0.219) **0.43**	67 (-0.191, -0.219) **0.276**	36 (0.578, -0.540) **0.654**	67 (-0.191, -0.219) **0.928**	67 (-0.191, -0.219) **0.884**	67 (-0.191, -0.219) **0.757**
15%	67 (-0.191, -0.219) **0.942**	67 (-0.191, -0.219) **0.44**	67 (-0.191, -0.219) **0.276**	50 (-0.139, 0.903) **0.677**	67 (-0.191, -0.219) **0.931**	67 (-0.191, -0.219) **0.874**	67 (-0.191, -0.219) **0.729**
20%	67 (-0.191, -0.219) **0.956**	82 (-0.233, -0.239) **0.42**	67 (-0.191, -0.219) **0.272**	2 (-0.161, -0.291) **0.678**	67 (-0.191, -0.219) **0.899**	67 (-0.191, -0.219) **0.894**	67 (-0.191, -0.219) **0.716**
25%	67 (-0.191, -0.219) **0.981**	67 (-0.191, -0.219) **0.43**	67 (-0.191, -0.219) **0.277**	20 (0.223, -0.127) **0.661**	67 (-0.191, -0.219) **0.941**	67 (-0.191, -0.219) **0.933**	67 (-0.191, -0.219) **0.715**

. – Observation number**; (.) –** Location**; Bold –** Depth value

It is observed that, "Mahalanobis, Halfspace, Simplicial, Spatial, Zonoid, Projection depths" tolerates upto 25% amount of contaminations and gives the same deepest point (measure of location).

3.3.3 Location and Scale Contamination

Case 5:

In this section results have been generated based on simulation study. For this study, the data were simulated (n=50) from normal distribution, mean vector μ= (0, 0), and unit covariance matrix, $\sum = I_2$. The various level of contaminations (mean vector, μ=(4,4) and unit covariance matrix, $\sum = 1.5I_2)$ such as 0%, 1%, 2%, 5%, 10%, 15%, 20% and 25% are considered and the obtained results are summarized given below.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

Table 3.7: Measure of location and the associated depth value under various data depth procedures

Error	MD	HSD	SD	SVD	SPD	ZD	PRD
0%	24 (0.092, 0.022) **0.988**	24 (0.092, 0.022) **0.44**	24 (0.092, 0.022) **0.306**	18 (0.202, 0.177) **0.727**	24 (0.092, 0.022) **0.975**	24 (0.092, 0.022) **0.939**	24 (0.092, 0.022) **0.793**
1%	24 (0.092, 0.022) **0.978**	24 (0.092, 0.022) **0.44**	24 (0.092, 0.022) **0.305**	39 (0.007, 0.433) **0.752**	24 (0.092, 0.022) **0.942**	24 (0.092, 0.022) **0.939**	24 (0.092, 0.022) **0.731**
2%	24 (0.092, 0.022) **0.970**	24 (0.092, 0.022) **0.44**	24 (0.092, 0.022) **0.306**	48 (0.434, 0.282) **0.742**	24 (0.092, 0.022) **0.966**	24 (0.092, 0.022) **0.915**	24 (0.092, 0.022) **0.776**
5%	24 (0.092, 0.022) **0.935**	24 (0.092, 0.022) **0.4**	24 (0.092, 0.022) **0.299**	24 (0.092, 0.022) **0.729**	24 (0.092, 0.022) **0.898**	24 (0.092, 0.022) **0.876**	24 (0.092, 0.022) **0.699**
10%	50 (0.513, 0.229) **0.970**	24 (0.092, 0.022) **0.38**	24 (0.092, 0.022) 0.299	25 (0.749, 0.578) **0.731**	24 (0.092, 0.022) **0.868**	50 (0.513, 0.229) **0.867**	24 (0.092, 0.022) **0.644**
15%	50 (0.513, 0.229) **0.989**	45 (0.528, 0.319) **0.32**	45 (0.528, 0.319) **0.283**	18 (0.202, 0.177) **0.739**	50 (0.513, 0.229) **0.829**	50 (0.513, 0.229) **0.958**	38 (0.007, 0.433) **0.570**
20%	50 (0.513, 0.229) **0.981**	50 (0.513, 0.229) **0.36**	50 (0.513, 0.229) **0.299**	50 (0.513, 0.229) **0.743**	50 (0.513, 0.229) **0.926**	50 (0.513, 0.229) **0.945**	24 (0.092, 0.022) **0.613**
25%	3 (0.997, 1.107) **0.999**	50 (0.513, 0.229) **0.380**	50 (0.513, 0.229) **0.296**	50 (0.513, 0.229) **0.732**	50 (0.513, 0.229) **0.930**	3 (0.997, 1.107) **0.985**	50 (0.513, 0.229) **0.652**

. – Observation number**; (.) –** Location**; Bold –** Depth value

It is observed that, “Mahalanobis and zonoid depth” tolerates upto 5%, “halfspace, simplicial, spatial and projection depth” tolerates upto 10% of contaminations. Simplicial volume depth does not performs well even if low level of contaminations.

Case 6:

This section presents the results of a simulation study. For this study, the data were simulated (n=100) from normal distribution, mean vector μ= (0, 0), and unit covariance matrix, $\Sigma = I_2$. The various level of contaminations (mean vector, μ=(4,4) and unit covariance matrix, $\Sigma = 1.5I_2)$ such as 0%, 1%, 2%, 5%, 10%,15%, 20% and 25% are considered and the obtained results are summarized in the following table.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

Table 3.8: Measure of location and the associated depth value under various data depth procedures

Error	MD	HSD	SD	SVD	SPD	ZD	PD
0%	57 (0.025, 0.027) **0.995**	57 (0.025, 0.027) **0.4**	39 (0.144, -0.118) **0.274**	91 (-0.070, 0.431) **0.687**	39 (0.144, -0.118) **0.916**	57 (0.026, 0.027) **0.956**	39 (0.144, -0.118) **0.756**
1%	48 (0.596, 0.119) **0.999**	68 (0.689, -0.955) **0.41**	39 0.144, -0.118) **0.274**	80 (-0.012, -0.375) **0.761**	39 (0.144, -0.118) **0.921**	48 (0.596, 0.119) **0.992**	68 (0.689, -0.956) **0.761**
2%	57 (0.025, 0.027) **0.996**	57 (0.025, 0.027) **0.42**	39 (0.144, -0.118) **0.276**	28 (-0.054, 0.250) **0.720**	39 (0.144, -0.118) **0.933**	57 (0.025, 0.027) **0.981**	39 (0.144, -0.118) **0.775**
5%	35 (0.248, 0.065) **0.989**	39 (0.144, -0.118) **0.44**	39 (0.144, -0.118) **0.277**	36 (0.019, 0.257) **0.739**	39 (0.144, -0.118) **0.946**	35 (0.248, 0.065) **0.950**	39 (0.144, -0.118) **0.832**
10%	83 (0.779, 0.713) **0.974**	36 (0.248, 0.065) **0.42**	36 (0.248, 0.065) **0.276**	60 (0.495, 0.138) **0.754**	36 (0.248, 0.065) **0.947**	36 (0.248, 0.065) **0.905**	36 (0.248, 0.065) **0.812**
15%	26 (0.019, 0.257) **0.978**	57 (0.025, 0.027) **0.46**	57 (0.025, 0.027) **0.278**	4 (0.359, -0.011) **0.749**	57 (0.025, 0.027) **0.987**	57 (0.025, 0.027) **0.946**	57 (0.025, 0.027) **0.946**
20%	18 (0.727, 1.152) **0.973**	96 (-0.017, 0.162) **0.41**	36 (0.019, 0.257) **0.272**	35 (0.248, 0.065) **0.707**	36 (0.019, 0.257) **0.894**	83 (0.779, 0.713) **0.900**	36 (0.019, 0.257) **0.774**
25%	83 (0.779, 0.713) **0.759**	83 (0.779, 0.713) **0.6**	60 (0.495, 0.138) **0.265**	36 (0.019, 0.257) **0.774**	35 (0.248, 0.065) **0.895**	83 (0.779, 0.713) 0.936	68 (0.689, -0.956) **0.662**

. – Observation number**; (.) –** Location**; Bold –** Depth value

It should be noted that very deep, "Simplicial and Spatial" allow a certain amount of pollution and give the same score as deep (measure of location). Other systems do not support depth, and although data contamination is severe, they do not provide a very reliable depth point.

In summary, halfspace, simplicial, spatial and projection depth performs well in the context (i) location contaminations, (ii) scale contaminations and (iii) location and scale contaminations. Specifically, halfspace and projection depth equally performs well when compared to other depth procedures.

4. Application in Discriminant Analysis

The applicability of data depth procedures is explored through discriminate analysis using real data. This approach is compared to the calculation of misclassification probabilities.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

Case 1: (Two groups)

Description: The hemophilia data (Habemma et al. (1974)) (Appendix: A9) contains two measured variables (X_1 = log10 (AHF activity) and X_2 = log10 (AHV antigen)) on 75 women, belonging to two groups: n1=30 (normal group) and n2= 45 (obligatory carries). The 53rd observation is identified as outlier through distance-distance plot (figure 4.1). The Discriminant analysis was performed under various depth procedures under with and without outliers. The deepest points and misclassification probabilities are summarized in the table 4.1 and 4.2 respectively. The depth contours plots of discrimination under various procedures are presented in appendix (Appendix: A3 and A4).

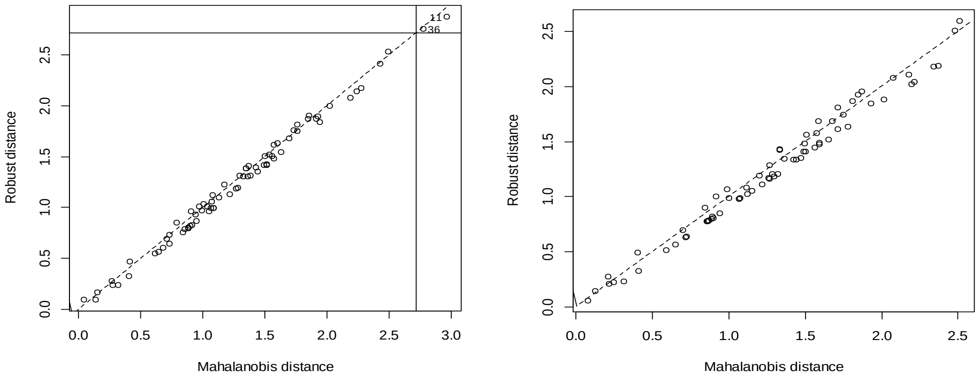


Figure 4.1: Distance-Distance Plots (with/without outliers) (hemophilia data)

Table 4.1: Measure of location and the associated depth value under various data depth procedures

Methods	MD	HSD	SD	SVD	SPD	ZD	PRD
With outlier	51 (-0.2447, -0.0407) **0.998097**	55 (-0.2154, -0.0219) **0.44**	55 (-0.2154, -0.0219) **0.286161**	55 (-0.2154, -0.0219) **0.670173**	55 (-0.2154, -0.0219) **0.941833**	51 (-0.2447, -0.0407) **0.977317**	55 (-0.2154, -0.0219) **0.782035**
Without outlier	55 (-0.2154, -0.0219) **0.993265**	55 (-0.2154, -0.0219) **0.438356**	55 (-0.2154, -0.0219) **0.288314**	20 (-0.2015, -0.0498) **0.686374**	55 (-0.2154, -0.0219) **0.960403**	55 (-0.2154, -0.0219) **0.950385**	55 (-0.2154, -0.0219) **0.778241**

. – Observation number**; (.) –** Location**; Bold –** Depth value

Table 4.2 Computed misclassification probabilities under various data depth procedures

Methods	MD	HSD	SD	SVD	SPD	ZD	PD
With outlier	0.2057	**0.1486**	0.1408	0.2394	0.1408	0.2057	**0.1486**
Without outlier	0.1507	**0.1268**	0.0986	0.2057	0.0986	0.1268	**0.1268**

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

From the table above it can be seen that HSD and PD have the same depth point when considering maximum scale depth with and without outliers. Each method works equally well with other methods. When comparing misclassification probabilities, all higher ratios performed better except for "Mahalanobis and the Simplicial data depth method".

Case 2: (Three groups)

Description: A real dataset is considered, namely the anorexia dataset (Hand et al. 1993) (Appendix: A12). The dataset consists of 3 groups, each group containing two variables with a base of 72 observations. Data on weight change in young anorexic patients. There are two variables, prewt (weight of patients before the study period) and postwt (weight of patients after the study period), classified into three groups, namely Cont (control), CBT (cognitive-behavioural therapy) and FT (family therapy). The 41st and 64th observations are identified as outlier through distance-distance plot (figure 4.2). The Discriminant analysis was performed under various depth procedures under with and without outliers. The deepest points and misclassification probabilities are summarized in the table 4.3 and 4.4 respectively. The depth contours plots of discrimination under various procedures are presented in appendix (Appendix: A5 and A6).

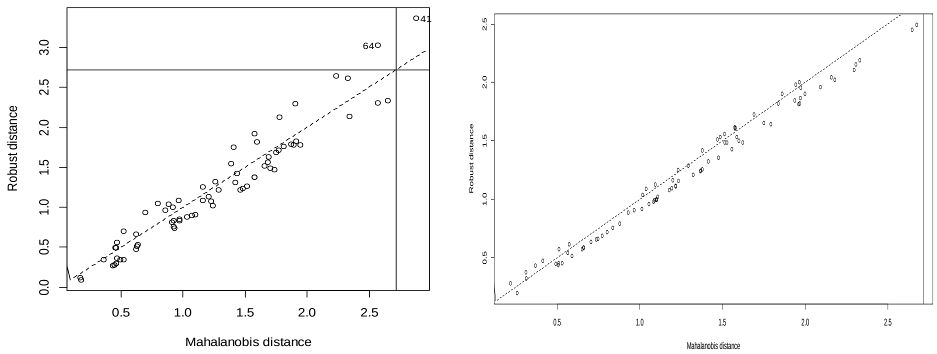


Figure 4.2: Distance-Distance Plots (with/without outliers) (anorexia data)

Table 4.3: Measure of location and the associated depth value under various data depth procedures

Methods	MD	HSD	SD	SVD	SPD	ZD	PRD
With outlier	43 (83.3, 85.4) **0.970367**	51 (83.3, 85.2) **0.388889**	51 (83.3, 85.2) **0.280818**	22 (84.4, 84.7) **0.685654**	51 (83.3, 85.2) **0.880519**	51 (83.3, 85.2) **0.91338**	51 (83.3, 85.2) **0.701694**
Without outlier(64)	51 (83.3, 85.2) **0.985547**	39 (81.3, 82.4) **0.4**	51 (83.3, 85.2) **0.283084**	29 (81.5, 81.4) **0.699589**	51 (83.3, 85.2) **0.886203**	39 (81.3, 82.4) **0.939106**	39 (81.3, 82.4) **0.700296**

. – Observation number**; (.) –** Location**; Bold –** Depth value

Table 4.4 Computed misclassification probabilities under various data depth procedures

Methods	**MD**	**HSD**	**SD**	**SVD**	**SPD**	**ZD**	**PD**
With outlier	0.4930	0.4930	**0.4507**	0.5352	**0.4507**	0.5070	0.5352
Without outlier	0.4853	0.4627	**0.4328**	0.4930	**0.4328**	0.4853	0.4507

From the above table the comparison of average probability of misclassification values in the above table, simplicial and spatial Depth performs better than the other methods. Since these two procedures gives low misclassification probabilities when compared with other data depth procedures.

In summary, “halfspace, projection, spatial and simplicial depth” provides low misclassification rate under with and without outliers when compared to other depth procedures such as “Mahalanobis, zonoid and simplicial volume depth”.

5. Conclusion

Local measurement is one of the most important concepts in statistical analysis. At this time, there is room for great information to be considered as a good metric for doing some analysis and for understanding the data. Over the past couple of years, many statistical methods have been advanced for estimating the spatial level, while the process of known depth is the newest method for determining a fixed location by observing the deepest data point in the cloud. In this context, this dissertation demonstrates the various concepts of information processing that have been introduced recently. To do this, he studied the situation by collecting real and simulated data in an environment. Moreover, the application of these processes to the most profound numerical studies has been carried out in the context of discrimination analysis.

Most widely used data depth procedures have been reviewed in this dissertation such as “Mahalanobis Depth, Half space Depth, Simplicial Depth, Simplicial Volume Depth, Zonid Depth and Spatial Depth”. The performance of these depth procedures has been studied under real data set and simulated environment. Among all depth procedures, halfspace and projection depth is recommended because of its remarkable properties, for example robustness, affine invariance, maximality at center, monotonicity relative to deepest point, vanishing at infinity, etc. Further it is noted that, though depth procedures work well in certain situations and in the context of their formulation, the depth procedures namely, halfspace, projection, simplicial and spatial depth performs more efficient than other discussed depth procedures. These procedures tolerate certain amount of abnormal observations in the data set. Further, in the context classification problems, these procedures give less misclassification error rate when compared with other depth procedures.

The study reveals that the half space and projection depth perform equally good and more efficient than other depth procedures. The research communities can get more accuracy while using these procedures in order to find the good location by identifying the deepest point in a data cloud, instead of using conventional measure of location. Since, measure of location and scale estimates find numerous applications to statistical inference and multivariate data analysis, data depth are geometric in nature, the study can be further explored in this context. Also, the future research may be carried out using robust statistics in data depth and vice versa, since the robust statistics and data depth are less influenced by abnormal observations. We can apply these proceduresin multivariatedata analysis techniques and helpful in the field of basic Sciencesresearch communities. Fortunately, computers with increasing processing power and larger memory is available now, which is good for the researcher and future of data depth.

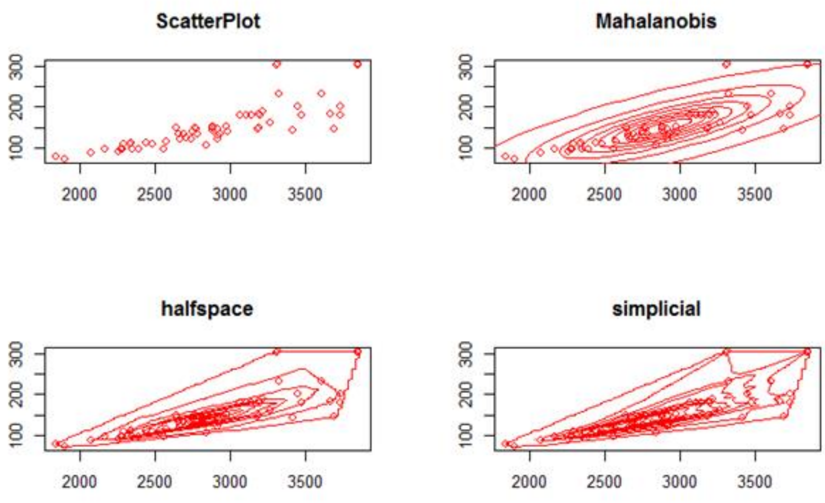
References

[1] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. Journal of the Americal Statistical Association 91 862–872.
[2] Dyckerhoff, R., Koshevoy, G., and Mosler, K. (1996). Zonoid data depth: theory and computa-tion. In: Prat A. (ed), COMPSTAT 1996. Proceedings in computational statistics, Physica-Verlag (Heidelberg), 235–240.
[3] Hubert, M.and VanDriessen,K.(2004). Fast and RobustDiscriminant Analysis. ComputaionalStatistics andData Analysis,45,301-320.
[4] Koshevoy, G. and Mosler, K. (1997). Zonoid trimming for multivariate distributions Annals of Statistics 25 1998–2017.
[5] Liu, R. Y. (1990). On a notion of data depth based on random simplices. The Annals of Statistics 18 405–414.
[6] Liu, R.Y. (1992). Data depth and multivariate rank tests. In: Dodge, Y. (ed.), L1-Statistics and Related Methods, North-Holland (Amsterdam), 279–294.

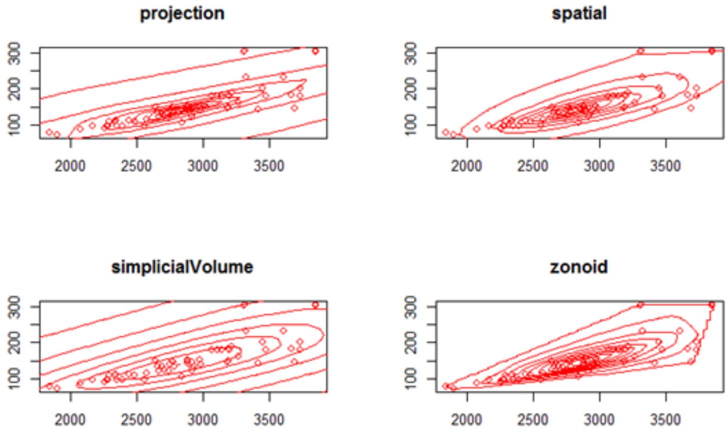
JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

[7] Liu, R.Y. J.M. Parelius and K. Singh (1999), Multivariate analysis by data depth: Descriptive Statistics, Graphics and Inference. The Annals of Statistics, 27, 783-858.
[8] Liu, X. and Zuo, Y. (2014). Computing projection depth and its associated estimators. Statistics and Computing 24 51–63.
[9] Mahalanobis, P. (1936). On the generalized distance in statistics. Proceedings of the National Academy India 12 49–55.
[10] Mahadevan, G. and Renuka, K. (2019), [1,2]-Complementary Connected Domination Number of Graphs-III, Commun. Fac. Sci. Univ. Ank. Ser. A1 Math. Stat, 68, 2298-2312.
[11] Maria Raquel Neto. The Concept of Depth In Statistics. Depth Article Paper.
[12] Oja, H. (1983). Descriptive statistics for multivariate distributions. Statistics & Probability Letters 1 327–332.
[13] Serfling, R. (2006). Depth functions in nonparametric multivariate inference. In: Liu, R., Serfling, R., Souvaine, D. (eds.), Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications, American Mathematical Society, 1–16.
[14] Subba Reddy, Ch. Yookesh, TL. and Boopathi Kumar, E. (2022), A Study On Convergence Analysis Of Runge-Kutta Fehlberg Method To Solve Fuzzy Delay Differential Equations. Journal of Algebraic Statistics, 13(2), 2832-2838.
[15] Tukey, J.W. (1975). Mathematics and the picturing of data. In: Proceeding of the International Congress of Mathematicians, Vancouver, 523–531.
[16] Vardi, Y. and Zhang, C. (2000), “The Multivariate L_1 Median and Associated Data Depth”, Proceedings of the National Academy of Science USA, 97, 1423-1426.
[17] Zuo, Y.J. and Serfling, R. (2000). General notions of statistical depth function. The Annals of Statistics 28 461–482.
[18] Zuo, y. (2003). Projection-based depth functions and associated medians, The Annals of statistics, 31, 1460-1490.

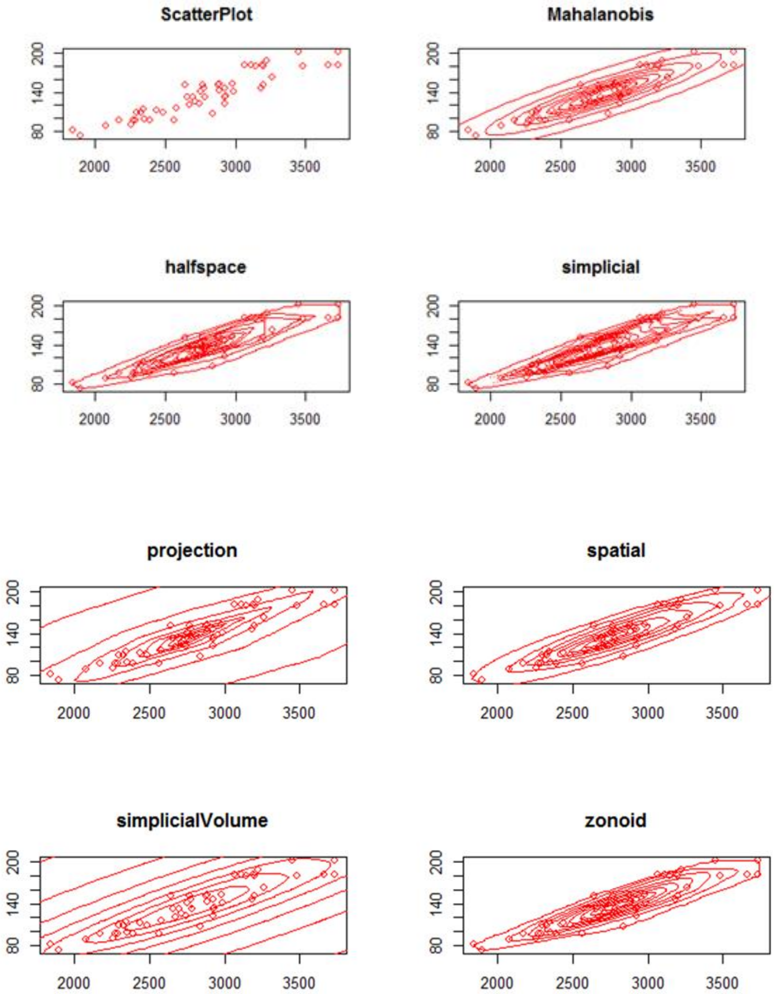
A1: Scatter Plot and Depth contoursunder various depth procedures (with outliers)(cardata90)



JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

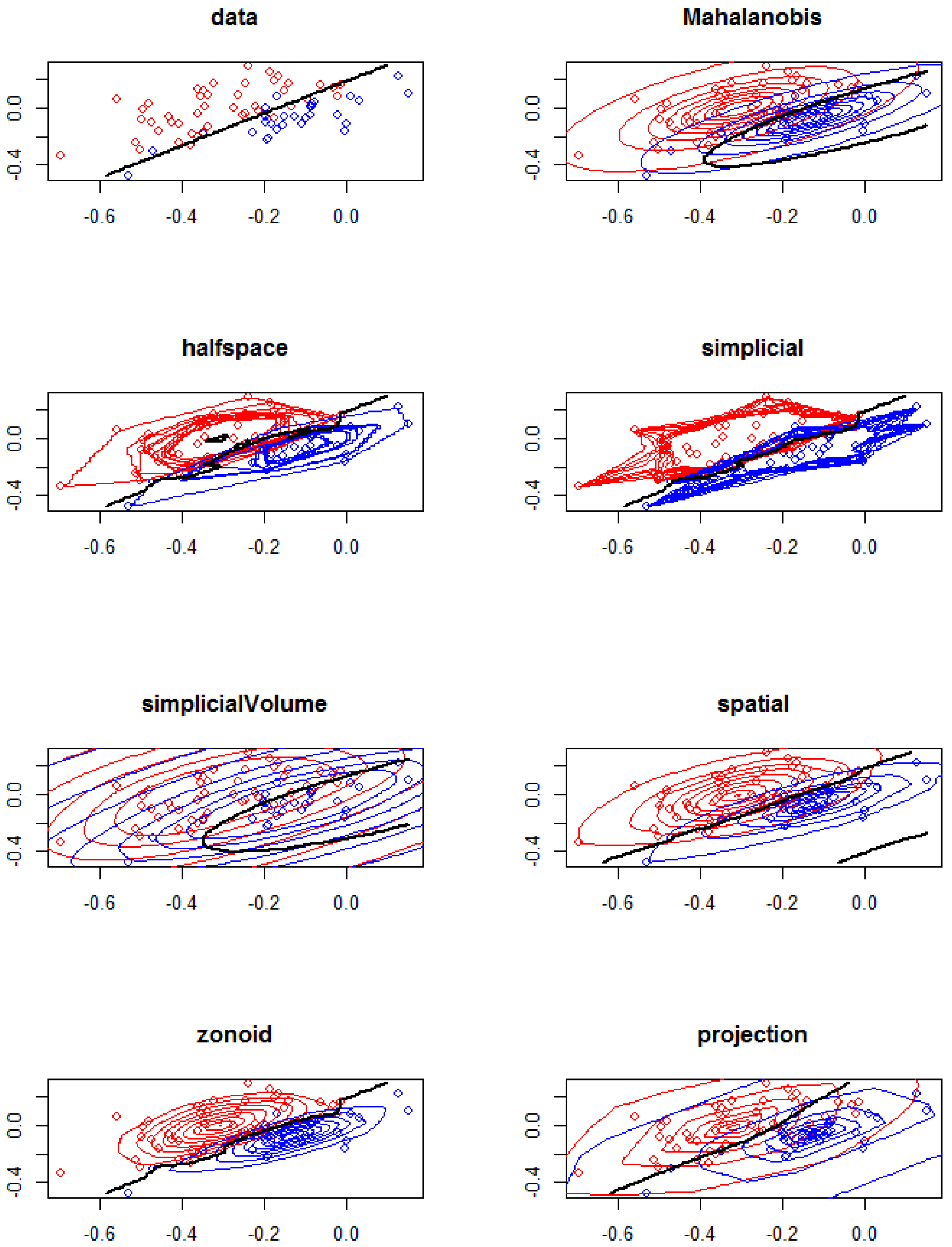


A2: Scatter Plot and Depth contoursunder various depth procedures (without outliers)(cardata90)



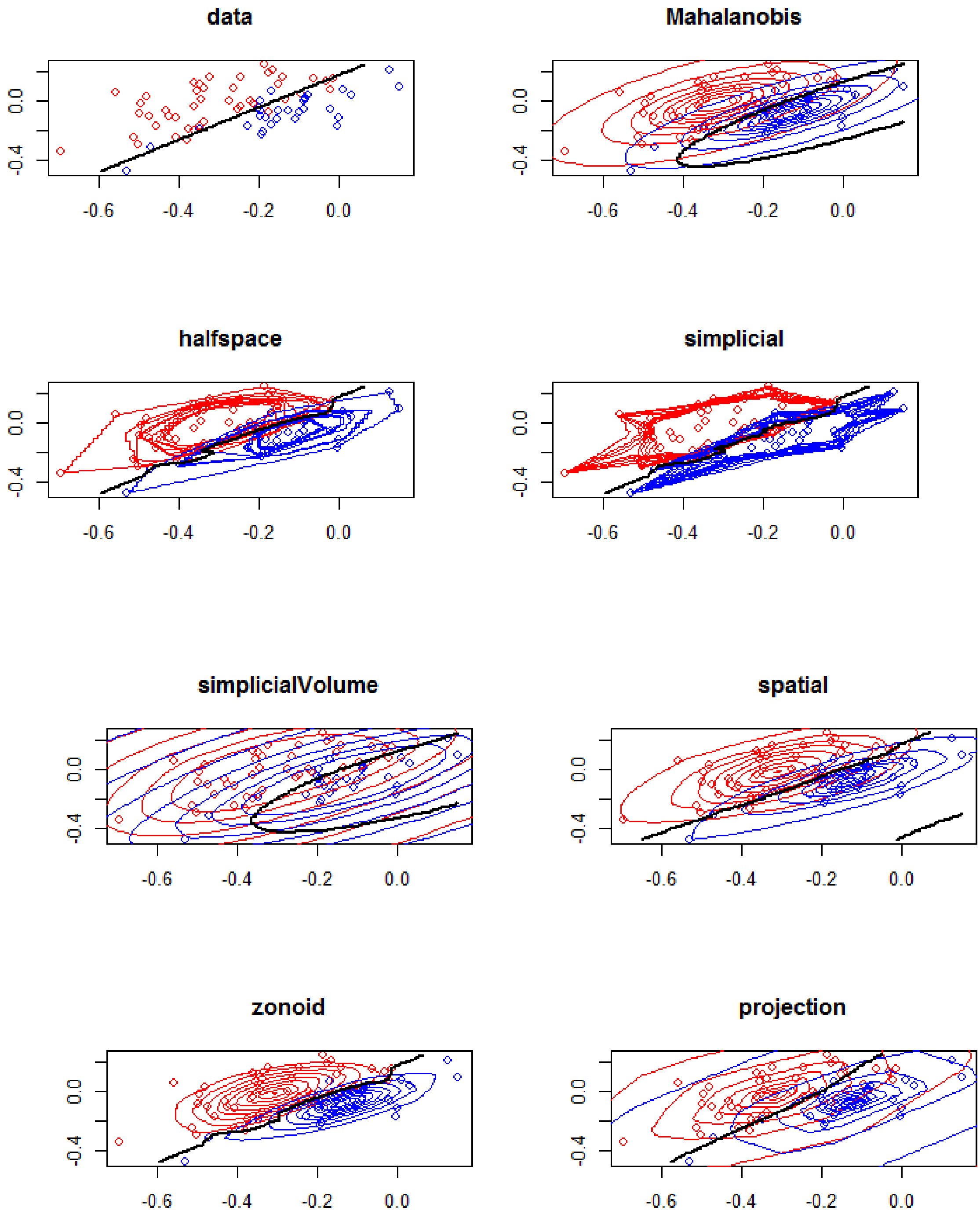
JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

A3: Depth contourswith classification under various depth procedures (with outliers) (hemophilia data)



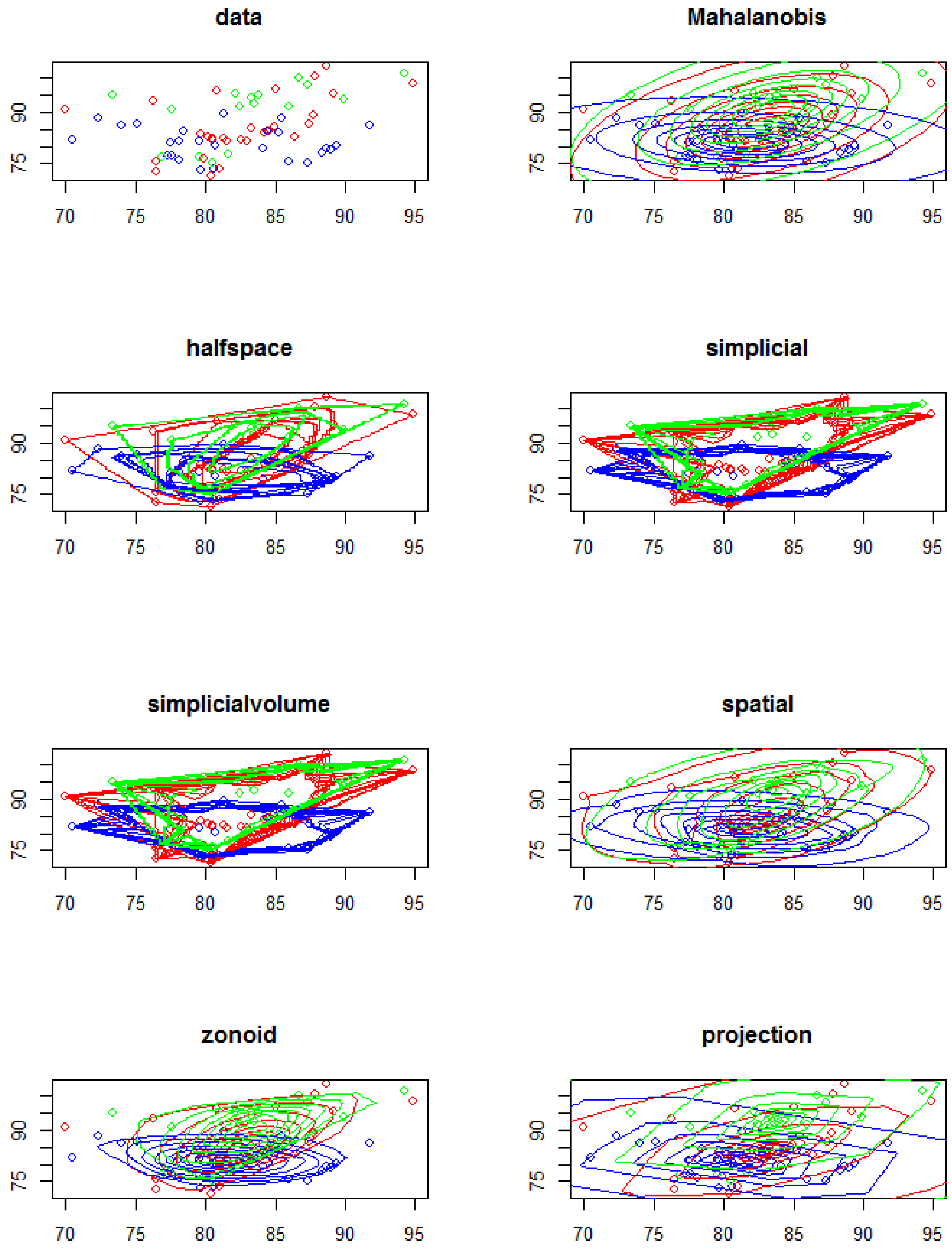
JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

A4: Depth contours with classification under various depth procedures (without outliers) (hemophilia data)



JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

A5: Depth contourswith classification under various depth procedures (with outliers) (anorexia data)



JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1995-2015
https://publishoa.com
ISSN: 1309-3452

A6: Depth contourswith classification under various depth procedures (without outliers)(anorexia data)