Journal of Al-Qadisiyah for Computer Science and Mathematics Vol. 14(4) 2022 , pp Stat.. 28–36

University Of AL-Qadisiyah

Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)

Some Methods for Detecting Outliers

***Hadeel Kamil Habeeb* [a] *Mohammed Al-Guraibawi* [b]**

[b] *Al-Qadisiyah University- College of Administration and Economics, Iraq . Email: hadeel_8080@gmail.com*

[b] *Al-Furat Al-Awsat Technical University- Diwaniyah Technical Institute, Iraq.Email: dw.moh2@atu.edu.iq*

ARTICLE INFO

Article history:
Received: dd /mm/year
Rrevised form: dd /mm/year
Accepted : dd /mm/year
Available online: dd /mm/year

https://doi.org/10.29304/jqcm.2022.14.4.1196

ABSTRACT

Analyzing data need detecting some outliers; therefore many important methods appeared and introduced to detect these outliers. In this paper, we consider the distinguished methods employed to detect the outliers. Our main purpose, shed light the classical methods for detecting outliers and fixing disadvantage of these methods.

MSC..

1. Introduction

There are many methods for detecting outliers in multivariate data. For example, leverage value, Mahalanobis distance. The classical Mahalanobis distance is a common method for detecting outliers. However, it is a method based on sample mean vector and sample covariance matrix. Since the classical mean vector and covariance matrix algorithms are sensitive to outliers, the classical Mahalanobis distance is also sensitive to outliers. Many authors have proposed robust estimation methods for mean vector and covariance matrix, such as *MVE* estimator, *MCD* estimator, etc.

The Mahalanobis distance was examined as a multivariate distance in this work, and its advantages over the Euclidean distance were discussed. It became clear that when dealing with correlated multivariate data, the Mahalanobis distance, which takes correlation into account, is preferable to the Euclidean distance. It was also shown that multivariate outliers can be detected using the Mahalanobis distances. To calculate the Mahalanobis distances, one must first estimate the theoretical mean vector and covariance matrix. Since outliers have a large impact on these estimators, estimating these parameters using their conventional empirical equivalents, especially when the data contain outliers, yields erroneous findings. One sensible solution is to use trustworthy statistical techniques.

A robust Mahalanobis distance can be calculated for each point using distance-based techniques like MCD, which are based on robust estimates of the mean and covariance matrix despite the existence of 593 different robust estimations. The aforementioned methods have been applied in this work to find multivariate outliers in a real data set using the statistical computing software environment R.

∗Corresponding author

Email addresses:

Communicated by ‘sub etitor’

In contemporary statistics, linear regression models represent a large and highly developed field. One of the most widely applied models in this field is the multiple linear regression model. The most common method used in multiple linear regression models is to establish a functional relationship between two or more quantitative variables so that one or more explanatory variables can predict a response variable. The significant topics in this chapter's literature reviews relate to robust diagnostic methods and robust estimation techniques for multiple linear regression models with outliers.

2. Depth Function and half space depth

Liu (1990) introduced an important new variety of depth function, the "simplicial depth", and emphasized the general role of a depth function as providing a center-outward ranking of data points. The one-dimensional case the spatial depth is equivalent to the halfspace depth, as seen from [30] as follow as:

$$D_s(\mathrm{x},\mathrm{F}) = 1 - |2F(x) - 1| = 2\min\{\, F(x), 1 - F(x)\}.$$

Regina Y. Liu (1990) introduced a new notion of data depth. This notion emerges naturally out of a fundamental concept underlying affine geometry, namely that of a simplex, and it satisfies the requirements one would expect from a notion of data depth. Thus it leads to an affine invariant, center-outward ranking of the data points. [32]

Donoho and Gasko (1992) explored the properties of the location depth and of the deepest location for finite data set, where the deepest location is a point with maximal ldepth which it the center of gravity the center of gravity of the innermost ldepth region. [9].

Rousseeuw and Ruts (1996) were the first to use the concept of a circular sequence to precisely calculate the half-space depth for bivariate data clouds and generate their outlines, where computed the depth of a point with complexity O(n log n), and a single depth region is constructed with complexity O(n^2log n), both essentially determined by the complexity of the QUICK-SORT procedure. [34]

Johnson et al. (1998) proposed accounting for only a small fraction of points while creating the first l depth contours, which results in a lower complexity for small l. (algorithm FDC). A data cloud is described by a finite number of depth contours in halfspace depth. [23]

Rousseeuw and Struyf (1998) introduce an algorithm to compute the halfspace depth for $d = 3$ with complexity O(n^2 log n). They use the Rousseeuw and Ruts (1999) procedure to determine the halfspace depth in these planes by projecting points onto planes orthogonal to the lines linking each of the points from X with z.[37].

2.1 **Regression depth**

Rousseeuw, Peter & Hubert, Mia (1999) propose the notion of regression depth. They view depth as a property of a fit (typically determined by a vector θ of coefficients), rather than a property of an observation. In general, they define the depth of a (candidate) fit θ to a given dataset Z_n of size n (θ, Z_n) to be the smallest number of observations of Z_n that would need to be removed to make θ a nonfat [35]. To compute the regression depth (rdepth) in O(n) operations using the following formula:

$$\mathrm{rdepth}(\theta, Z_n) = \min_{1 \le i \le n} (\min\{L^+(x_i) + R^-(x_i), R^+(x_i) + L^-(x_i)\}) \,,$$

where $L^+(v) = \#\,\{\mathrm{j}; x_j \le v \; and \; r_j \ge 0\}, R^-(v) = \#\,\{\mathrm{j}; x_j > v \; and \; r_j \le 0\}$, and L^- and R^+ are defined accordingly.

2.2 **The contaminants**

!Geoff Robson (2003) mention that the contaminants are outliers caused by human error or the presence of a separate generation mechanism and a different distribution. Outlying observations are typically not contaminants except in the case of heavy-tailed distributions such as Student's t. The contamination of samples drawn from the normal distribution, which is not prone to outliers, was addressed. It is also assumed that nothing about the distribution's parameters is known a priori, which is usually the case.[33].

Miller et al. (2003) compute halfspace depth for bivariate data clouds, and the halfspace depth describes a data cloud by a finite number of depth contours with complexity $O(n^2)$, and the depth of a single point may then be computed with complexity ($\log^2 n$)[29]. Bremner et al. (2006) use a primal-dual technique to determine the halfspace depth by incrementally updating the upper and lower boundaries using a heuristic until they coincide [3].

In [10], functional depths—a measure of a curve's centrality inside a collection of trajectories that provides center outward orderings of the set of curves—are used to assess outlier detection for functional data. They offer some explanations regarding the value of searching for outliers in functional datasets and suggest an in-depth strategy for functional outlier detection. Numerous Monte Carlo trials are used to examine the performance of the suggested approach. We conclude by locating outliers in a dataset of NOx (nitrogen oxides) emissions obtained from a control station close to an industrial region to demonstrate the approach.

Moreover, Bremner et al. (2008) design an output-sensitive depth-calculating algorithm that represents the task as two maximum subsystem problems for $d > 2$ [4]. In order to visualize functional data and spot functional outliers, the researchers in [21] suggest some new tools. The suggested tools leverage high-density regions, deep data, and robust principal component analysis. We demonstrate that our approaches outperform the existing "functional depth" method in identifying outliers in French male age-specific mortality data by comparing the proposed outlier detection methods with it.

Kernelized spatial depth (KSD) (2009), a statistical depth function, and an outlier identification technique were presented by Yixin Chen [9]. The spatial depth has been generalized by the KSD where the form of KSD can be explain as

$$\mathrm{D}_{\kappa}(x,\chi) = 1 - \frac{1}{|\chi - \{x\}| - 1} \times \left(\sum_{y,z\,\in \chi} \frac{\kappa(x,x) + \kappa(y,z) - \kappa(x,y) - \kappa(x,z)}{\delta_{\kappa}(x,y)\,\delta_{\kappa}(x,z)} \right)^{1/2},$$

where $\delta_{\kappa}(x,y) = \sqrt{\kappa(x,x) + \kappa(y,y) - 2\kappa(x,z)}$.

In a feature space caused by a positive definite kernel, it defines a depth function. Any observation's KSD can be assessed using a specific set of samples. In a feature space caused by a positive definite kernel, it defines a depth function. Any observation's KSD can be assessed using a specific set of samples. As a data point moves further from the center, or the spatial median, of the data cloud, the depth value, which is always within the interval [0, 1]. This inspires a straightforward algorithm for detecting outliers, which labels an observation as one if its KSD value is below a predetermined threshold. We arrived at the probabilistic inequality for the likelihood that an outlier detector will generate false alarms. The threshold of an outlier detector can be selected to regulate the upper bound on the probability of false alarms under a specific level using these inequalities. They tested the suggested technique for outlier detection on both simulated and real-world data sets. On all test data sets, the KSD based outlier detection performs competitively.

3. Multivariate Functional and halfspace depth

The zonoid depth was pioneered by Mosler et al. (2009). They took advantage of the notion to divide R^d into direction cones, and later techniques for determining depth and depth regions, including the halfspace depth, did the same [30]. Mosler consider the definition of zonoid regions as $\boldsymbol{D_0(F) = R^d}$ and for $\boldsymbol{\alpha \in [0,1]}$

$$\boldsymbol{D_{\alpha}(F) = \left\{ \int_{R^d} x g(x) dF(x) : 0 \le g \le \frac{1}{\alpha}, \int_{R^d} g(x) dF(x) = 1 \right\}}.$$

A direct link between multivariate quantile areas and halfspace depth trimmed regions is shown by Hallin et al. (2010). [17]

When bivariate depth and depth lines continually add points to the data set, updating depth becomes a fascinating problem, which Burr et al. (2011) explores. [6]

W. S. Lok & Stephen M.S. Lee (2011) proposed a new statistical depth function based on interpoint distances, which has the distinct property of respecting multimodality in data configurations, which it proves to be especially relevant to many inference problems including confidence region construction, classification, tests for equality of populations, p-value computation, etc. With specification of an appropriate interpoint distance, our depth function also applies to infinite-dimensional data. Where the conventional centre-outward ordering depth functions are found to be inadequate. [26]

In order to demonstrate that their envelope coincides with the appropriate halfspace depth trimmed region, Kong and Mizera (2012) use direction quantiles, which are halfspaces that correspond to quantiles on univariate projections. For $d > 2$, the areas of depth are precisely computed [24].

Ieva and Paganoni (2013) used Multivariate Functional Principal Component Analysis to reduce the dimensionality of their data. It entails multiplying the respective scores by the information contained in the covariance operators of the signals and their first derivatives. Projecting data and derivatives onto the relevant Karhunen-Loève bases yields scores. [4]

Liu and Zuo (2014) use a breadth-first search technique to cover R^d and QHULL to define the direction cones in order to precisely determine the halfspace depth. For the precise computation of the halfspace depth, he offers two more, seemingly quick procedures. This algorithm is one of them; it is called a refined combinatorial algorithm [25].

Sara López-Pintado et al. (2014) proposed Simple depth-of-range concepts for multivariate functional data that extend univariate functional depth of range, providing simple and natural criteria for measuring path centrality within a sample of curves.

Recall that the standard simplicial depth $\boldsymbol{SD(y; P_Y)}$ of a multivariate vector y in R^d with respect to the multivariate distribution $\boldsymbol{P_Y}$ is defined as

$$\boldsymbol{SD(y; P_Y) = P\{y \in simplex\{Y_1, \dots, Y_{p+1}\}}.$$

Based on these depths, a sample of multivariate curves can be ordered from the center out and system statistics can be determined. The proposed depths have characteristics, such as stability and consistency [27].

Ieva, Francesca et al. (2015) developed statistical methods to compare two independent samples of multivariate functional data that differ in terms of covariance factors. The concept of depth measurement has been generalized to this type of data, taking advantage of the role of contrast factors in weighting the components that determine depth. It was applied to electrocardiogram (ECG) signals aimed at comparing physiological subjects and affected patients with left bundle branch block. Also, the proposed depth scales calculated on the data were used to perform a non-parametric comparison test between these two groups. They are also presented in a generalized regression model that aims to classify ECG signals.[22]

4. Clustering and Tukey depth

Katie Evans et al. (2015) devised a method to identify outlying observations in model-based clustering based on normal mixture models that influence cluster structure and number, without identifying clusters amid a wide range of noisy observations. The outliers are those with a minimum membership proportion or for which the cluster-specific variance with and without the observation is very different. The method demonstrated its ability to detect true outliers without incorrectly identifying many non-outliers and improves performance compared to other approaches [13].

!Reyes, Alicia & Cuesta-Albertos, Juan (2015) proposed a modification of the first procedure in Hubert et al. (2015) consisting in basing it on the random Tukey depth, where the random Tukey depth is a statistical depth that approximates the Tukey depth. It needs of a very low number of projections to obtain equivalent results to those of the Tukey depth. So, the random Tukey depth is very fast to compute, making it the depth to go for, not only when the dimension of the space is moderate or high, but also when it is low due to its computationally effectiveness. Additionally, the random Tukey depth inherits from the Tukey depth the nice properties that made it well known. Also he proposed a simpler and more usual measure of variation [19].

Rainer Dyckerhoff, Pavlo Mozharovskyi (2016) proposed a theoretical framework for computing the halfspace depth, which yields a whole class of algorithms. The data for each of these tuple is projected onto the corresponding orthogonal complement, and the halfspace depth was computed as the sum of the depth in these two orthogonal subspaces and all proposed algorithms are capable of dealing with data that is not in general mode and even with ties.[31]

5. Mahalanobis distances

According to Olusola Samuel (2017), is an alternative to several parametric methodologies in evaluating large amounts of multivariate data. A nonparametric classification strategy based on several data depth function conceptions is addressed, and certain features of these approaches are investigated. The performance of various depth functions in maximum depth classifiers is explored using simulation and real data in the agriculture business. [28]

!Dutta, Subhajit & Genton, Marc (2017) Use depth-based estimates to construct regression estimates, and investigate their performance with respect to existing estimators. To increase the efficiency of this estimator, a re-weighted estimator based on strong Mahalanobis distances from the remaining vectors has been proposed. The method is more stable than current methods that are generated using subsamples of data from an empirical point of view. The resulting multivariate regression technique is computationally feasible, and has been shown to perform better than many popular robust multivariate regression methods when applied to diverse simulation data as well as a real reference dataset. When the dimension of the data is too high compared to the sample size, meaningful concepts of data depth can still be used along with corresponding depth values to create a robust estimator in a sparse environment.[17]

Mia Hubert et al. (2017) created classifications of multivariate and functional data in order to combine novel stability, robustness, and computational feasibility. On the basis of the halfspace depth, the bag distance (BD) has been proposed. It meets the majority of the features of a norm and can also represent asymmetry. Instead of delving into the facts. In addition, a DistSpace transformation based on bd or an outlyingness metric is proposed, followed by k-nearest neighbor (kNN) categorization of the changed data points. This combines kNN's wide applicability and endurance with its stability and simplicity. The concept was tested against other approaches using actual and simulated data.[20]

!Taban Baghfalaki and Mojtaba Ganjali (2017) proposed a robust generalized estimating equations (RGEE) that based on statistical depth and extend the approach to robust weighted generalized estimating equations (RWGEE), which express centrality of points with respect to the whole sample with a smaller depth (larger depth) for the point far from the center (for the point near the center)[2] . Harsh, Archit et al. (2018) present and implement a modified onion peeling algorithm to detect top-k outliers in a Gaussian 2-D data set. The idea of onion peeling, or peeling in short, is to construct a convex hull around all the points in the dataset and then find the points that are on the edge of the convex hull. These points form the first "peel" and are removed from the dataset. Repeating the same process gives more and more peels, each containing a number of points. We modified this basic idea to detect the k largest outliers in a given 2-D Gaussian data-set. The choice of k is influenced by the spatial geometry of the data-set and is user-defined. The convex hull is the smallest convex set that contains all of the points in the set.[18]

5.1 MAHALANOBIS DISTANCE AND ITS APPLICATION FOR DETECTING MULTIVARIATE OUTLIERS

The paper [15] reviewed the Mahalanobis distance as a multivariate distance and discussed its benefits over the Euclidean distance. It was made obvious that the Mahalanobis distance, which accounts for correlation, is preferable to the Euclidean distance when working with correlated multivariate data. Additionally, it was demonstrated how the Mahalanobis distances can be used to spot multivariate outliers. The researcher consider the Mahalanobis distance definition as :

$$D(X,\mu) = \sqrt{(X-\mu)^T \Sigma^{-1} (X-\mu)}\,.$$

If the covariance matrix is the identity matrix, the Mahalanobis distance reduces to the Euclidean distanceThe theoretical mean vector and covariance matrix must be estimated in order to compute the Mahalanobis distances. Since outliers, estimating these parameters using their typical empirical counterparts, significantly impact these estimators particularly when data contain outliers, produces false results. Using reliable statistical methods is one sensible solution. Although there are 593 various robust estimates, distance-based techniques like MCD are based on robust estimates of the mean and covariance matrix, allowing for the computation of a robust Mahalanobis distance for each point. With the help of the R software environment for statistical computing, the aforementioned techniques have been used in this study to identify multivariate outliers in a real data set.

5.2 Mahalanobis Distance and Multivariate Outlier Detection in R

A useful distance metric for determining the separation between a point and a distribution is the Mahalanobis distance (MD) see [8]. On data with multiple variables, it works fairly well. MD leverages covariance across variables to determine the separation between two locations, which is why it is successful on multivariate data. In other words, Mahalanobis determines the separation between points ''P1'' and ''P2'' by taking standard deviation into account (how many standard deviation P1 far from P2). Even when outliers are treated as multivariate data, MD still produces trustworthy results. Every point's distance from the center in N- dimensional data is determined in order to locate outliers via MD, and taking these distances into account discovers outliers.

5.3 Multivariate Outlier Detection by Using Two- Dimensional Correlation

This study's application of the two-dimensional correlation approach to identify outliers in precipitation data with just 12 observed values was successful (months in a year) [10]. The pairwise approach employed in comparisons of the observations rather than taking into account the distributions as a whole makes it possible for the method to have this attribute. This benefit might make it possible for researchers to find outliers that are missed by more traditional techniques. The suggested approach can be used to identify outliers in a variety of other research fields because it has a generic application that extends beyond hydrology. Future research may focus on adapting the concept to areas where outliers are important, where the outliers have a big impact where more "regular" data don't. By comparing seasonal series, such as yearly comparisons of records from a hydrologic station, the approach may also be used to identify univariate outliers. The 2DCorrel software created to apply the methodology is publicly offered as an addition to make it simple for researchers to reimplement the strategy.

5.4 Outlier detection in multivariate functional data through a contaminated mixture model

The activity of sensors is frequently monitored in an industrial setting. Automatically detecting anomalous measurement behavior is difficult [1]. The issue can be stated as the identification of outliers in a multivariate functional data set by treating the sensor measurements as functional data. The suggested contaminated mixture model detects outliers and clusters the multivariate functional data into homogenous groups as a result of the heterogeneity of the data set. The key benefit of this method over others is that the percentage of outliers is not have to be specified. The BIC is used to choose the number of clusters, and the Expectation-Conditional Maximization technique is utilized to accomplish model inference. Numerical tests using simulated data show the inference technique to have a high level of performance. Particularly, the proposed model performs better than the rivals. Its use with the real data that drove this study makes it possible to accurately identify anomalous behavior.

5.5 Robust Mahalanobis distances

!Elisa Cabana, Rosa E. Lillo and Henry Laniado (2021) proposed a set of robust Mahalanobis distances based on the concept of shrinking to detect multivariate outliers. Shrinkage is best determined by estimating robust intensities and scaling factors. And some properties were investigated, including equation value and hash. When we deviate from the common assumption of normality, the behavior in a simulation and a real data set shows the appropriateness of the method. The advantages of our proposal have been demonstrated by the resulting high correct detection rates and low false detection rates in a large number of cases, as well as significantly shorter computation time.[7]

González-De La Fuente et al. (2022) studied a statistical data depth with respect to compact convex random sets, which is consistent with the multivariate Tukey depth and the Tukey depth for fuzzy sets. In addition, it provides a different perspective to the existing halfspace depth with respect to compact convex random sets. They provided a series of properties for the statistical data depth with respect to compact convex random sets. These properties are an adaptation of properties that constitute the axiomatic notions of multivariate, functional, and fuzzy depth-functions and other well-known properties of depth. [16]

6. Conclusion

In order to the role of the methods to detect the outliers with different formula wither was related to the halfspace depth or Mahalanobis distance, we consider in this work some of the methods related with. The Mahalanobis distance was examined as a multivariate distance in this work, and its advantages over the Euclidean distance were discussed. It became clear that when dealing with correlated multivariate data, the Mahalanobis distance, which takes correlation into account, is preferable to the Euclidean distance. It was also shown that multivariate outliers can be detected using the Mahalanobis distances.

References

[1] *M. Amovin-Assagba*, *I. Gannaz* and *J. Jacques*, "Outlier detection in multivariate functional data through a contaminated mixture model ", *Computational Statistics & Data Analysis*, vol. 174, issue C, 2022.

[2] Baghfalaki, T. & Ganjali, M. (2017). Robust Weighted Generalized Estimating Equations Based on Statistical Depth. Communications in Statistics - Simulation and Computation. 46. 10.1080/03610918.2016.1277746.

[3] Berrendero, J.R. & Justel, A. & Svarc, M. (2011). Principal components for multivariate functional data. Computational Statistics & Data Analysis. 55. 2619-2634. 10.1016/j.csda.2011.03.011.

[4] Bremner, D. & Fukuda, K. & Rosta, V. (2006). Primal-dual algorithms for data depth. 10.1090/dimacs/072/12.

[5] Bremner, D. & Chen, D. & Iacono, J. & Langerman, S. & Morin, P. (2008). Output-sensitive algorithms for Tukey depth and related problems. Statistics and Computing. 18. 259-266. 10.1007/s11222-008-9054-2.

[6] Burr, M. & Rafalin, E. & Souvaine, D. (2011). Dynamic Maintenance of Half-Space Depth for Points and Contours.

[7] Cabana, E., Lillo, R.E. & Laniado, H. (2021). Multivariate outlier detection based on a robust Mahalanobis distance with shrinkage estimators. Stat Papers 62, 1583–1609. https://doi.org/10.1007/s00362-019-01148-1

[8] Cansiz S., "Mahalanobis Distance and Multivariate Outlier Detection in r."2020.

[9] Y. Chen, X. Dang, H. Peng and H. L. Bart, "Outlier detection with the Kernelized spatial depth function," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, pp. 288–305, Feb. 2009.

[10] F. Dikbas, "Multivariate Outlier Detection by Using Two-Dimensional Correlation" , CC BY 4.0 License, 2021.

[11] Donoho, D. & Gasko, M. (1992). Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness. Ann. Stat.. 20. 10.1214/aos/1176348890.

[12] Dutta, S. & Genton, M. (2017). Depth-weighted robust multivariate regression with application to sparse data. Canadian Journal of Statistics. 45. 10.1002/cjs.11315.

[13] Evans K, Love T, Thurston SW. (2015). "Outlier Identification in Model-Based Cluster Analysis". J Classif. 32(1):63-84. doi: 10.1007/s00357-015-9171-5.

[14] Febrero M., Galeano P. and Gonz´alez-Manteiga W., "Outlier detection in functional data by depth measures, withapplication to identify abnormal NOx levels "Wiley InterScience, Environmetrics ,19(4):331-345, 2007.

[15] Ghorbani H. , " MAHALANOBIS DISTANCE AND ITS APPLICATION FOR DETECTING MULTIVARIATE OUTLIERS ", Ser. Math. Inform. Vol. 34, No 3 (2019), 583–595 .

[16] González-De La Fuente, L., Nieto-Reyes, A., Terán, P. (2022). Properties of Statistical Depth with Respect to Compact Convex Random Sets: The Tukey Depth. Mathematics, 10, 2758. https://doi.org/10.3390/ math10152758.

[17] Hallin, M. & Paindaveine, D. & Siman, M. (2010). Multivariate Quantiles and Multiple-Output Regression Quantiles: From L1 Optimization to Halfspace Depth. The Annals of Statistics. 38. 635-669. 10.1214/09-AOS723.

[18] Harsh, A., Ball JE, Wei P. (2018). "Onion-Peeling Outlier Detection in 2-D data Sets.", International Journal of Computer Applications, 139(3)pp. 26-31.

[19] M. Hubert, P. Rousseeuw and P. Segaert: Multivariate functional outlier detection. Statistical Methods & Applications. **24**, PP. 237–243, (2015)

10.1007/s10260-015-0319-6.

[20] Hubert, M. & Rousseeuw, P. & Segaert, P. (2017). Multivariate and functional classification using depth and distance. Advances in Data Analysis and Classification. 11. 445–466. 10.1007/s11634-016-0269-3.

[21] R. Hyndman and H. Lin Shang,"Bagplots, Boxplots and Outlier Detection for Functional Data". Functional and Operatorial Statistics. Contributions to Statistics. Physica-Verlag HD, 2008.

[22] Ieva, F. & Tarabelloni, N. & Paganoni, A. & Biasi, R. (2015). Use of Depth Measure for Multivariate Functional Data in Disease Prediction: An Application to Electrocardiograph Signals. The International Journal of Biostatistics. 11. 10.1515/ijb-2014-0041.

[23] Johnson, T., Kwok, I., and Ng, R. (1998). Fast computation of 2- dimensional depth contours. In: Agrawal, R., and Stolorz, P. (eds.), Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, AAAI Press, New York, 224–228.

[24] Kong, L. & Mizera, I., Quantile tomography: Using quantiles with multivariate data. Statistica Sinica. Vol. 22, No. 4 (October 2012), pp. 1589-1610. 10.5705/ss.2010.224.

[25] Liu, X. & Zuo, Y. (2014). Computing Halfspace Depth and Regression Depth. Communications in Statistics: Simulation and Computation. 43. 10.1080/03610918.2012.720744.

[26] Lok, W. S. & Lee S. M. (2011) A new statistical depth function with applications to multimodal data, Journal of Nonparametric Statistics, 23:3, 617-631, DOI: 10.1080/10485252.2011.553953.

[27] Lopez-Pintado, S. & Sun, Y. & Lin, J. & Genton, M. (2014). Simplicial band depth for multivariate functional data. Advances in Data Analysis and Classification. 8. 321-338. 10.1007/s11634-014-0166-6.

[28] Makinde, O. S. & Adewumi, A. D. (2017). A comparison of depth functions in maximal depth classification rules. Journal of Modern Applied Statistical Methods, 16(1), 388-405. doi: 10.22237/jmasm/1493598120.

[29] Miller, K. & Ramaswami, S. & Rousseeuw, P. & Sellarès, J. & Souvaine, D. & Streinu, I. & Struyf, A. (2003). Efficient computation of location depth contours by methods of computational geometry. Statistics and Computing. 13. 153-162. 10.1023/A:1023208625954.

[30] Mosler, K., Lange, T., and Bazovkin, P. (2009), “Computing zonoid trimmed regions of dimension d > 2”, Computational Statistics and Data Analysis 53, 2500-2510.

[31] Rainer D., Pavlo M.. (2016). Exact computation of the halfspace depth, Computational Statistics & Data Analysis, Volume 98, PP. 19-30, ISSN 0167-9473, https://doi.org/10.1016/j.csda.2015.12.011.

[32] Regina Y. Liu (1990). "On a Notion of Data Depth Based on Random Simplices," The Annals of Statistics, Ann. Statist. 18(1), 405-414.

[33] Robson, G. (2003). “Multiple outlier detection and cluster analysis of multivariate normal data.”.

[34] Rousseeuw, P. & Ruts, I. (1996). Algorithm AS 307: Bivariate Location Depth. Journal of the Royal Statistical Society Series C Applied Statistics. 45. 516-526. 10.2307/2986073.

[35] Rousseeuw, P. & Hubert, M. (1999). Regression Depth. Journal of the American Statistical Association. 94. 388-402. 10.1080/01621459.1999.10474129.

[36] Serfling, R. (2002). A Depth Function and a Scale Curve Based on Spatial Quantiles.

[37] Struyf, A. & Rousseeuw, P. (1999). Halfspace Depth and Regression Depth Characterize the Empirical Distribution. Journal of Multivariate Analysis. 69. 135-153. 10.1006/jmva.1998.1804.