doi: 10.21608/esju.2024.306371.1037

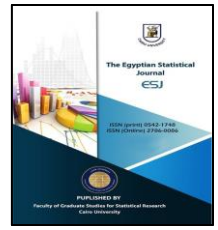
Research Article

The Egyptian Statistical Journal (ESJ), 68 (2): 24-33

Homepage: https://esju.journals.ekb.eg/

The Egyptian Statistical Journal

Print ISSN 0542-1748– Online ISSN 2786-0086



Review: Detecting Outliers with Distributions for Estimating Time Series Models

Mohamed Ezzat Abd-Elaziz [1*], Ahmed Amin El-Sheikh[2], Amal Mohamed Abd-Elfatah [2]

Received 22 July 2024; revised 25 September 2024; accepted 1 October 2024

Keywords

Outlier, Time series, ARIMA, Autoregressive, Exponential, Gamma, Additive outlier.

Abstract

In statistical analysis, outliers represent data points that significantly deviate from the general pattern of a dataset. Understanding and addressing outliers is crucial because they can skew results, impacting the reliability and validity of conclusions drawn from data analysis. This paper provides an in-depth exploration of outliers across three key dimensions. First, it offers a general overview of outliers, discussing their characteristics, methods, and algorithms used to detect them in various contexts, including fields such as finance, healthcare, and social sciences. Second, it examines outliers within distributions, detailing how they influence measures such as mean, median, variance, and standard deviation, and how they can affect the overall shape and interpretation of the data distribution. Techniques for detecting and mitigating the impact of outliers are also discussed. Third, it analyzes outliers within time series data, focusing on their potential to distort trends, cyclic patterns, and forecasting accuracy. By investigating these dimensions, this paper aims to enhance the understanding of outliers and underscore their significance and challenges in statistical analysis.

1. Introduction

An outlier refers to a data point that stands out either due to its notably small or large value or because it diverges from the overall pattern observed in the dataset. The topic of outliers has garnered considerable attention in recent decades, indicating that the exploration of outliers is not a novel concept. The majority of real-world datasets exhibit outliers characterized by values that significantly differ from others within the dataset, potentially impacting the accuracy of data analysis. Various approaches and examinations exist for identifying outliers, tailored to the nature of the data, including methods and tests for cross-sectional data adhering to specific distributions, as well as methods and tests designed for time series data. Numerous statisticians have investigated techniques for identifying outliers, which may arise in either cross-sectional or time series data. These studies can be categorized based on the nature of the data under examination.

This paper aims to elucidate the methodologies and perspectives delineated in prior research endeavors. It is structured as follows: Section (2) Review of Outliers Problem, Section (3)

* Corresponding author*: Mohamed.ezzat.abdelazez@gmail.com

[1] Faculty of Management, Economics and Business Technology, Egyptian Russian University.

[2] Faculty of Graduate Studies for Statistical Research, Cairo University.

Detecting of Outliers in Distributions, Section (4) Detecting of Outliers in Time Series Models and Section (5) Conclusion.

2. Review of Detection Methods for Outliers

This section will present other methods for detecting outliers in data sets regardless of distribution.

Jiang and An (2008) introduced a clustering-based outlier detection method (CBOD) and expanded the notion of an object's outlier factor to clusters. The outlier factor of the cluster (i) $OF(C_i)$ quantifies the degree to which a cluster i is an outlier; a higher value indicates a greater likelihood of the cluster being an outlier. In this approach, clusters generated through the clustering process are treated as individual units and classified as either "normal" or "outlier" clusters. The method can detect one or multiple outliers and consists of two stages: first, the dataset is grouped using a one-pass clustering algorithm; second, the resulting clusters are evaluated and labeled as "normal clusters" or "outlier clusters" based on their outlier factor. The process for determining outlier clusters involves the following steps:

1. Compute outlier factor $OF(C_i)$, where $1 \leq i \leq k$.
2. Sort clusters $C = \{C_1, C_2, \ldots, C_k\}$ and make them satisfy:
$$OF(C_1) \geq OF(C_2) \geq \cdots \geq OF(C_k)$$
3. Search the minimum b which satisfies $\frac{\sum_{i=1}^{b}|C_i|}{|D|} \geq \epsilon(0 < \epsilon < 1)$, where $|C_i|$ and $|D|$ are the number of observations in the cluster C_i and the number of all observations in the dataset respectively, and ϵ is an approximate ratio of the outlier to the whole training dataset, if there is no information about the dataset ϵ is selected in the range of $[0.05, 0.1]$
4. The clusters $C_1, C_2, .., C_b$ are considered outliers and any observation belongs these clusters is an outlier, the clusters $C_{b+1}, C_{b+2}, \ldots, C_k$ are considered as normal clusters and any observation belongs these clusters considered as not outlier.

A Thorough performance study assessed the proposed method using real-world datasets from the UCI Machine Learning Repository. The datasets included the Lymphography dataset (1988), which consists of 148 records with 18 attributes, and the Wisconsin Breast Cancer dataset (1992), which has 699 records with 9 numerical attributes. The performance of the outlier detection methods was measured using the detection rate (DR) and false alarm rate (FR). To validate the effectiveness of the proposed method, it was compared with the cluster-based local outlier factor (FindCBLOF) method by He et al. (2003) and the tensor-based outlier detection (TOD) method by Jiang et al. (2005). The findings revealed that the proposed method outperformed both the TOD and FindCBLOF methods.

Al-Zoubi et al. (2010), suggested a method for identifying outliers, capable of detecting one or more outliers using fuzzy clustering techniques. Initially, the c-means algorithm is applied to produce an objective function. Small clusters are then identified and considered as outlier clusters, defined as clusters with fewer points than half the average number of points in the c clusters.

To detect the outliers in the remaining clusters the following parameters are used: Objective function (OF) for the entire set, (OF_i) the objective function after removing point x_i from the set, T = (1.5) and (DOF_i) difference between OF and OF_i Following steps are conducted:

1. Run the Fuzzy c-means (FCM) algorithm to generate an Objective Function (OF).
2. Identify small clusters and classify the points within these clusters as outliers.
3. For the remaining points (those not identified as outliers in Step 2), perform the following calculations for each point x_i, calculate the following: OF_i, DOF_i, and average DOF
4. If ($DOF_i > T$(average DOF) then classify x_i as an outlier.

The objective function represents the (Euclidean) sum of squared distances between cluster centers and the points assigned to those clusters, weighted by the membership values generated by the c-means algorithm. The method was tested on three real datasets which are, the Wood dataset (20 points, six dimensions), the Bupa dataset (six dimensions) and the Iris dataset (four dimensions), all from the UCI repository. The performance of the proposed method was compared to the Clustering-Based (CB) method proposed by Al-Zoubi (2009). Using the detection rate as the performance metric, the results indicated that the proposed method was more effective at detecting outliers than the CB method.

Kannan and Manoj (2015) conducted a comparison of various distance measures which are Mahalanobis Distance, Cook's Distance, Leverage, and Difference in Fits, to detect one or more outliers in multivariate data. Traditional outlier detection methods based on the sample mean and covariance matrix often fail to produce optimal results because they are influenced by the presence of outliers. They used a diabetes dataset with 80 observations and 8 variables (Age, Pregnancy, Plasma, Pressure level, Skin cells, Insulin level, Body Mass Index (BMI), and Pediatric). Distances calculated using the four measures were plotted on a scatter plot to identify outliers. The results showed that the outlier detection capabilities of Mahalanobis Distance and Leverage Point were roughly equivalent, while Difference in Fits had very low sensitivity for detecting outliers. Cook's Distance, however, exhibited very high sensitivity, identifying the maximum number of outlier points. This indicates that Cook's Distance is particularly effective at identifying the most highly affected diabetes patients.

Abuzaid (2020), enhanced the Local Outlier Factor (LOF) method which was proposed by Breunig et al. (2000) to deal with the problem of outliers in multivariate circular data, it was regarded as the pioneering method for addressing the challenge of outlier detection in multivariate circular data. The (LOF) for a point (i) is calculated by determining its average distance to its k nearest neighbors. A higher LOF value indicates sparser neighborhoods, suggesting an outlier point, whereas a lower LOF value signifies denser neighborhoods, indicating a normal point. The performance of the extended test was evaluated against other numerical test statistics which are: (M) test statistic proposed by Mardia (1975), (C) test statistic proposed by Collett (1980), (D) test statistic proposed by Collett (1980), (A) test statistic proposed by Abuzaid et al. (2009) and (G) test statistic proposed by Mohamed et al. (2016). The null hypothesis and alternative hypothesis of these test statistics are formulated as follows:

H_0: observation x_i is not outlier and H_1: x_i is outlier where $1 \leq i \leq n$.

Three measures were used for evaluating the performance which are the probability of type-II error (P1), the probability of incorrectly classifying a contaminated value as an outlier when it is actually an extreme value (P3), and the probability of wrongly identifying a good observation as discordant (P1-P3). A simulation study evaluated the extended method's performance for multivariate circular data by generating 2000 random samples from two circular distributions: the von Mises distribution (mean μ, concentration parameter ρ) and the wrapped Cauchy distribution (mean ν, concentration parameter υ). The results indicated that the extended method was compatible with the (A) test and outperformed other tests.

3. Detecting of Outliers in Distributions

Zerbet and Nikulin (2003), proposed a test statistic called (Z_k) to detect upper outliers in exponential distribution. A sample of size (n) was assumed to be independent random variables. The null hypothesis and slippage alternative hypothesis were formulated as follows:

H_0: sample derived from an exponential distribution with parameter θ.
H_k: the first (n-k) order statistics derived from an exponential distribution with parameter θ and the rest (k) order statistics derived from an exponential distribution with parameter $\frac{\theta}{a}$, where $0 < a < 1$, a is unknown and k called upper outliers.
The distribution of the test based on the proposed test statistic was determined. Accordingly, tables of critical values were provided for various sample sizes (n) and numbers of outliers (k). A simulation study was conducted to compare the power of the proposed test statistic and the power of Dixon's statistic (D_k) (Chikkagoudar and Kunchur, 1983) for $\alpha = 0.05\ and\ 0.1$, $k = 3$, sample size varying from 6 to 12 and a varying from $0.01\ to\ 1$. The test utilizing the new statistic demonstrates greater power compared to the test relying on Dixon's statistic.
Nooghabi et al. (2010) extended the test statistic (Z_k) which was proposed by Zerbet and Nikulin (2003) to (Z_k^*) for the detection of upper outliers in gamma distribution. It was assumed that the sample size (n) consisted of independent random variables. The null hypothesis and slippage alternative hypothesis were formulated as follows:
H_0: sample derived from Gamma distribution with parameters $m\ and\ \gamma$.
H_k: the first (n-k) order statistics derived from a Gamma distribution with parameter $m\ and\ \gamma$, and the rest (k) order statistics derived from a Gamma distribution with parameter $m\ and\ \frac{\gamma}{d}$, where $d > 1$, d is unknown and k called upper outliers. The distribution of the test based on this statistic was determined, and tables of critical values were provided for various sample sizes (n) and numbers of upper outliers (k). The power of the extended test was also calculated and compared to the power of Dixon's statistic D_k through a simulation for $\alpha = 0.05\ and\ 0.1$, $k = 1,2,3$ and sample size varying from 6 to 11and d varying from $1.05\ to\ 2$. Results showed that the extended test Z_k^* was more powerful than the test based on Dixon's statistic and the critical value of the extended test Z_k^* increased as the sample size (n) increased. However, the critical value of D_k decreased as the sample size (n) increased. Also, the critical value of the extended test Z_k^* decreased when k increased. However, the critical value of D_k increased as the number of outliers k increased.
Lalitha and Kumar (2012) proposed two test statistics L_1 and L_k for testing single outlier and multiple outliers respectively in data that follows Exponential distribution. Tables for critical values are unnecessary as they can be easily calculated for any sample size. Null and slippage alternative hypotheses for L_k test statistic are formulated as follows:

H_0: $X_1, X_2, \dots, X_n$ are derived from an Exponential distribution with parameter θ.

H_k: $X_{(1)}, X_{(2)}, \dots, X_{(n-k)}$ are derived from exponential distribution with parameter θ, but $X_{(n-k+1)}, \dots, X_{(n)}$ observations are derived from exponential with parameter θc, where $c > 1$ and $X_{(1)} X_{(2)}, \dots, X_{(n)}$ denote the order statistics corresponding to the observations $X_1, X_2, \dots, X_n$. Null hypothesis and slippage alternative hypothesis for L_1 test statistic are special cases from L_k test statistic when $k = 1$. The performance of the proposed tests was compared with some existing tests like D_k and Z_k which were proposed by Likeš (1967) and Zerbet and Nikulin (2003) through some measures of performance like non-spurious power, spurious power and swamping effect. A simulation study was performed by generating Exponential samples with a unit scale across different sample sizes ($n = 20, 50, 120$). The study also considered the presence of outliers, with k set to 2 and 3, across 10,000 replications, and varied the parameter c from 1.5 to 50. The results pointed out that the proposed test performed better than the tests Z_k and D_k for any k and n in terms of power and non-spurious power. Additionally, as n and k increased, the differences in swamping effects between L_k and Z_k, as well as between L_k and D_k decreased. The proposed test had another advantage that the critical values could be obtained easily without the required tables.

Nooghabi (2017), extended two test statistics (B_k) and (Z_k) which were proposed by Basu (1965) and Zerbet and Nikulin (2003) respectively to $\bar{B}_k$ and $\bar{Z}_k$ for detecting outliers in Exponentiated Pareto distribution. These statistics were adaptations of methods for detecting outliers in Gamma and Exponential distributions. He assumed that $x_1, x_2, \dots x_n$ were arbitrary independent random variables from Exponentiated Pareto distribution and wanted to test if the sample contained outliers or not. The null hypothesis and slippage alternative hypothesis were assumed as follows:

H_0: $x_1, x_2, \dots x_n$ are derived from an Exponentiated Pareto distribution with parameters τ and υ.

H_k: the first $(n-k)$ order statistics are derived from an Exponentiated Pareto distribution with parameters τ and υ, while the remaining k order statistics are derived from an Exponentiated Pareto with parameters τp and υ, where $p > 1$, p is unknown. If the null hypothesis was rejected that means that the sample contains outliers. The distribution of the test based on these statistics was determined, and tables of critical values were provided for various sample sizes (n) and numbers of upper outliers (k) such that $k \leq \frac{n}{2}$. He compared the power of the extended test statistics through a simulation study for levels of significance $\alpha = 0.05\ and\ 0.1$ and identified that ($\bar{Z}_k$) was more powerful than the test based on ($\bar{B}_k$) for all values of n and k, then he described an example from an insurance company and the extended test statistic $\bar{Z}_k$ identified the outlier at 0.05 and 0.1 level of significance.

Jäntschi (2019), proposed a test statistic for the detection of outliers in uniform U(0,1) distribution. The proposed test statistic can be applied to any continuous distribution to detect outliers by constructing a confidence interval for the extreme value in the sample. This interval is determined at a certain (preselected) risk of being in error and depends on the sample size. The proposed statistic applies to known distributions and is also dependent on the statistical parameters of the population. The maximum likelihood estimation was used to determine the parameters of the uniform distribution. A simulation study was carried out, and two distinct strategies were devised to effectively manage large datasets. The results showed that the proposed test statistic performs better than the classical Grubbs' test in detecting outliers.

Deiri (2021), extended the (Z_k) test statistic proposed by Zerbet and Nikulin (2003) to (N_k^*) test statistic for the detection of outliers in Rayleigh distribution and compared the results with the generalized Dixon's statistic (D_k). The null and slippage alternative hypotheses were formulated as follows:

H_0: sample derived from Rayleigh distribution with parameter ω where $\omega > 0$.

H_k: the first $(n-k)$ order statistics are from Rayleigh distribution with parameter ω but (k) observations are from Rayleigh distribution with parameter ωh where $h > 1$, n is the sample size and k is the upper outliers. The distribution of the extended test statistic was determined, and tables of critical values were provided for various sample sizes n and number of outlier $k = 1,2,3$ and level of significance $\alpha = 0.05\ and\ 0.10$. The power of the extended test (N_k^*) was computed and compared with the power of Dixon's statistic (D_k) through a simulation for $\alpha = 0.05\ and\ 0.1$, $k\ = 1,2,3$ and sample size varying from 10 to 25. The results showed that the extended test (N_k^*) was more powerful than the test based on Dixon's statistic, additionally, the critical value of (N_k^*) increased when n increased while the critical value of (D_k) decreased when n increased, but the critical value of (N_k^*) decreased when k increased. But, the critical value of (D_k) increased when k increased.

4. Detecting of Outliers in Time Series Models

Zaharim et al. (2009), proposed a test statistic (η) to detect single additive outlier (AO) in ARMA (1, 1) models. Two simulation studies were conducted to investigate the properties of (η) concerning different sample sizes and to evaluate the detection performance of the proposed test statistic. The first simulation was conducted to investigate the properties of (η) concerning different sample sizes (n), and different values for ARMA (1, 1) coefficients, for each possible combination of sample size and coefficients values of ARMA (1, 1) 500 model were generated. The second simulation was conducted for the detection performance of (η), the same sample size and coefficient values of the ARMA (1, 1) model in the first simulation were used in addition to the AO effect (ω) of different magnitudes. For each possible combination of sample size, coefficients value of ARMA (1, 1) and magnitudes of AO effect (ω) 500 AO-contaminated series were generated by introducing an AO of the respected (ω) at T = n/2. Results from the first simulation study indicated that the test statistic increased as the sample size increased. In the second simulation study, the performance of the test statistic improved with larger magnitudes of the outlier effect. Results indicated that there was no apparent correlation between the selection of coefficients for ARMA (1, 1) and either the sampling behavior or the detection performance of η.

Louni (2008), extended the sequential test statistic (T) proposed by Abraham and Yatawara (1988) for detecting and identifying the outlier, and named it a modified sequential test (T^*). The extended test statistic effectively detects and categorizes outliers as either additive outliers (AO) or innovative outliers (IO) simultaneously and cohesively. The null and alternative hypotheses for the extended test statistic were formulated as follows:

H_0: There is no outlier

H_1: There is a single outlier.

After identifying the position of the outlier, the decision regarding whether it is an additive outlier (AO) or an innovative outlier (IO) is based on the sign of the (S) statistic, as proposed by Abraham and Yatawara (1988). If the (S) statistic is positive, the outlier is classified as an AO; conversely, if it is negative, the outlier is classified as an IO. Upon detecting an outlier and determining its type, its impact is estimated and subsequently removed from the residuals or the observations based on its type. Following this adjustment, the model parameters are recalculated using the corrected data series. This iterative process continues until no significant outliers remain in the data. A simulation study for comparison between the extended test statistic (T^*) and test statistic (T) was conducted by generating 100 observations according to the model AR(1). The series contained either a single additive outlier or a single innovational outlier at the location $t = 50$. The results indicated that the modified sequential test (T^*) outperformed the sequential test (T) particularly when dealing with innovational outliers (IO).

Ahmar et al. (2018), proposed a method called $ARIMA$ additive outlier method $(ARIMA - AO)$ for detecting and correcting data which contain additive outliers depending on the iterative procedure which proposed by Chang et al. (1988). By using this method $ARIMA$ model was obtained to fit the data containing additive outliers (AO). The steps of the proposed method are:

1. Estimating the parameter of $ARIMA$ model assuming that no outliers in the data using least square regression equation and fit the model.
2. Calculating the residuals of the fitted model.
3. Calculating $(\lambda_{A,T})$ test statistic proposed by Chang et al. (1988), for the presence of AO for the residual, the null and alternative hypotheses of this test statistic were formulated as follows:
   H_0: no AO at time t.
   H_1: there is AO at time t.
4. If H_0 was rejected or there was AO at time t, the effect of AO would be removed from the residual and recalculated the $(\lambda_{A,T})$ test statistic.

The previous steps are repeated till all outliers are identified. Adjusting the observations in a time series affected by outliers is achieved through the construction of generalized estimating equations. A simulation study using R application was conducted, an $AR(2)$ model was obtained. The model parameters were significantly different from zero, but the residuals were not normal due to the presence of outliers. The iterative procedure was applied till the residual became stationary, three outliers were detected and the new $AR(2)$ model was obtained. Results showed that there was an improvement in the Mean Square Error (MSE) value of 47.34% of the initial model.

Laome et al. (2021), developed $ARIMA$ additive outlier method $(ARIMA\ AO)$ which was proposed by Ahmar et al. (2018). This method forecasts time series data containing outliers by estimating the magnitude of the outlier and adjusting the original series accordingly. Through iterative procedures, $ARIMA\ AO$ effectively mitigates the impact of outliers, enhancing the accuracy of forecasts. The steps of the developed method are:

1. Estimating the parameter of $ARIMA$ model assuming that no outliers in the data.
2. Calculating the residual from the estimated model.
3. Calculating ($\lambda_{A,T}$) test statistic proposed by Chang et al. (1988), for the presence of AO, null and alternative hypotheses of this test statistic were formulated as follows:
   H_0: no AO at time t.
   H_1: there is AO at time t.
4. If H_0 was rejected or there was AO at time t, the original series would be modified by subtracting the AO effect from it, then the new residual would be calculated and also ($\lambda_{A,T}$) test statistic.

The previous steps are repeated till all outliers are identified. The method was applied to real data. The ARIMA AO model was implemented after obtaining some initial ARIMA models. Based on the analysis conducted using the ARIMA method for the data, it was assumed that the appropriate $ARIMA$ was $ARIMA$ $(0,1,1)$ with MSE and Mean absolute percentage error (MAPE) respectively 569140 and 10.05%. Whereas with the ($ARIMA\ AO$) method MSE and MAPE were 242544 and 7.37%. These results showed that the ($ARIMA\ AO$) method had a greater forecasting accuracy than the $ARIMA$ method.

5. Conclusion

This paper has provided a comprehensive exploration of outliers, shedding light on their critical role in statistical analysis. By examining outliers across three distinct dimensions, namely a general overview, distributions, and time series data, this study has highlighted the multifaceted nature of outliers and their potential impact on data analysis. The review of outliers in distributions focused exclusively on continuous distributions, indicating an area for further exploration in discrete distributions. Additionally, the examination of outliers in general revealed that most methods or algorithms rely on distance measures, such as Cook's and Mahalanobis distances, suggesting potential for alternative approaches. Furthermore, in the context of time series, the review concentrated predominantly on IO (Innovative Outliers) and AO (Additive Outliers), highlighting an opportunity to investigate other types of outliers. Through this investigation, it has become evident that understanding outliers is paramount for ensuring the accuracy, reliability, and validity of conclusions drawn from statistical analyses. Moving forward, continued research and the development of robust methodologies for detecting and managing outliers will be essential for enhancing the integrity of data-driven decision-making processes across various fields and industries.

Declaration of interests

The authors declare that they have no conflict of interest.

References

Abuzaid, A. H. (2020). Detection of outliers in univariate circular data by means of the outlier local factor (LOF). *Statistics in Transition. New Series*, 21(3): 39-51. https://doi.org/10.21307/stattrans-2020-043

Abuzaid, A. H., Mohamed, I. B., and Hussin, A. G. (2009). A new test of discordancy in circular data. *Communications in Statistics-Simulation and Computation*, 38(4): 682-691. https://doi.org/10.1080/03610910802627048

Ahmar, A. S., Guritno, S., Rahman, A., Minggi, I., Tiro, M. A., Aidid, M. K., and Arifin, A. N. M. (2018). Modeling data containing outliers using ARIMA additive outlier (ARIMA-

AO). *Journal of Physics: Conference Series*, 954(1): 012010. IOP Publishing. iopscience.iop.org/article/10.1088/1742-6596/954/1/012010.

Al-Zoubi, M. B. (2009). An effective clustering-based approach for outlier detection. *European Journal of Scientific Research*, 28(2): 310-316.

Al-Zoubi, M. B., Al-Dahoud, A., and Yahya, A. A. (2010). New outlier detection method based on fuzzy clustering. *WSEAS Transactions on Information Science and Applications*, 7(5): 681-690.

Breunig, M. M., Kriegel, H. P., Ng, R. T., and Sander, J. (2000). LOF: Identifying density-based local outliers. International conference on Management of data: 93-104. dl.acm.org/doi/pdf/10.1145/342009.335388

Chang, I., Tiao, G. C., and Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics*, 30(2): 193-204. https://doi.org/10.1080/00401706.1988.10488367

Chikkagoudar, M. S., and Kunchur, S. H. (1983). Distributions of test statistics for multiple outliers in exponential samples. *Communications in Statistics-theory and Methods*, 12(18): 2127-2142. https://doi.org/10.1080/03610928308828596

Collett, D. (1980). Outliers in circular data. *Journal of the Royal Statistical*, 29(1): 50-57. https://doi.org/10.2307/2346410

Deiri, E. (2021). A new statistic for detecting outliers in Rayligh distribution. *Life Cycle Reliability and Safety Engineering*, 10(2): 135-138. https://doi.org/10.1007/s41872-020-00150-z

Jäntschi, L. (2019). A test detecting the outliers for continuous distributions based on the cumulative distribution function of the data being tested. *Symmetry*, 11(6): 835. https://doi.org/10.3390/sym11060835

Jiang, S. Y., and An, Q. B. (2008). Clustering-based outlier detection method. International conference on fuzzy systems and knowledge discovery: 429-433. ieeexplore.ieee.org/abstract/document/4666153.

Kannan, K. S., and Manoj, K. (2015). Outlier detection in multivariate data. *Applied mathematical sciences*, 9(47): 2317-2324.

Lalitha, S. and Kumar, N. (2012). Multiple outlier test for upper outliers in an exponential sample. *Journal of Applied Statistics*, 39(6): 1323-1330. https://doi.org/10.1080/02664763.2011.645158

Laome, L., Wibawa, G. N. A., Raya, R., Makkulau, and Asbahuna, A. R. (2021). Forecasting time series data containing outliers with the ARIMA additive outlier method. *Journal of Physics: Conference Series*, 1899(1): 012106. IOP Publishing. iopscience.iop.org/article/10.1088/1742-6596/1899/1/012106/meta

Likeš, J. (1967). Distribution of Dixon's statistics in the case of an exponential population. *Metrika*, 11(1): 46-54. https://doi.org/10.1007/BF02613574

Louni, H. (2008). Outlier detection in ARMA models. *Journal of time series analysis*, 29(6): 1057-1065. doi:10.1111/j.1467-9892.2008.00595.x

Mardia, K. V. (1975). Statistics of directional data. *Journal of the Royal Statistical*, 37(3): 349-371. https://www.jstor.org/stable/2984782

Mohamed, I. B., Rambli, A., Khaliddin, N., & Ibrahim, A. I. N. (2016). A new discordancy test in circular data using spacings theory. *Communications in Statistics-Simulation and Computation*, 45(8): 2904-2916. https://doi.org/10.1080/03610918.2014.932799

Nooghabi, M. J. (2017). Detecting outliers in exponentiated pareto distribution. *Journal of Sciences, Islamic Republic of Iran*, 28(3): 267-272.

Nooghabi, M. J., Nooghabi, H. J., and Nasiri, P. (2010). Detecting outliers in gamma distribution. *Communications in Statistics-Theory and Methods*, 39(4): 698-706. https://doi.org/10.1080/03610920902783856

Zaharim, A., Rajali, R., Atok, R. M., Mohamed, I., and Jafar, K. (2009). A simulation study of additive outlier in ARMA (1, 1) model. *International Journal of Mathematical Models and Methods in Applied Sciences*, 3(2): 162-169.

Zerbet, A., and Nikulin, M. (2003). A new statistic for detecting outliers in exponential case, *Communications in Statistics-theory and Methods*, 32(3): 573-583. https://doi.org/10.1081/STA-120018552