Vol. 14. Issue.1. 2026 (Jan-March)

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

http://www.bomsr.com
Email:editorbomsr@gmail.com

RESEARCH ARTICLE

INTERNATIONAL STANDARD SERIAL NUMBER 2348-0580

Comparison of the Methods for detecting the Outliers

Ramnath Takiar
Scientist G (Retired), National Centre for Disease Informatics and Research (NCDIR), Indian Council of Medical Research (1978–2013), Bengaluru – 562110, Karnataka, India.
Email: ramnath_takiar@yahoo.co.in | ramnathtakiar@gmail.com
DOI:10.33329/bomsr.14.1.25


Ramnath Takiar

Article Info

Article Received: 10/01/2026
Article Accepted: 22/02/2026
Published online: 04/03/2026

Abstract

The Outliers in a series of data may correspond to extremely low or extremely high values and may result in a skewed distribution and may affect the normality of the data. In the presence of Outliers, the relationship studied between any two variables like income and expenditure may alter significantly. It is therefore essential to know how to identify the Outliers in a data set and how to treat them for further analysis.

For the present study, three method are used for the detection of the Outliers from the selected three normal samples of size 15, 20 and 25. According to selected three methods to identify the Outliers, the Lower Fence (LF) and Higher Fence values are defined as follows:

IQR-Old Method: LF = Q_1 – 1.5*IQR and HF= Q_3+ 1.5* IQR.

IQR-Takiar method: LF = Q_1 – IQR*[0.25*ln(n)+0.20] and HF= Q_3+ IQR*[0.25*ln(n)+0.20].

SD-Range-Takiar method: LF= Mean - SD*(0.37*ln (n) + 0.86) and

HF = Mean + SD*(0.37*ln (n) + 0.86). where n is the sample size.

None of the methods showed the presence of Outliers in the initially selected samples of size 15, 20 and 25. For each initial sample, the minimum and maximum values are replaced by still lower and higher values thereby defining the new 10 samples with Outliers. For the

study purposes, any value lying outside the range of each initial sample is assumed to be an Outlier. The Percentage detection rate of Outliers by the IQR-Old, SD-Range-Takiar method and IQR-Takiar method is 42%, 80% and 100%, respectively. The IQR-Takiar method is observed to be the superior method as compared to other two methods in detecting the Outliers and therefore, recommended to be used for identification of Outliers in the samples.

KEY WORDS: IQR-Old method, IQR-Takiar method, SD-Range-Takiar method, Outliers, Outlier detection rate.

1. Introduction

The Outliers in a series of data may correspond to extremely low or extremely high values. Their presence in a set of data often distorts the shape and pattern. In case of single variable, the presence of Outliers may often result in a skewed distribution and may affect the normality of the data. Outliers can severely inflate the mean, and particularly the variance, quite significantly. Sometimes, in the presence of Outliers, the relationship studied between any two variables like income and expenditure or age and mortality may present a distorted form of relationship between them. It is therefore essential to know how to identify the Outliers in a data set and how to treat them for further analysis. In literature, there are different methods available (Grubbs F.E.,1969, Schwertman NC, Owens MA, Adnan R 2004, STAT200/Elementary Statistics 2026, Takiar R 2023-1, Takiar R 2023-2, Vinutha HP., Poornima B, Sagar BM 2018).

By detection and appropriate management of outliers, one can obtain the better estimates of the Mean and SD giving the better representativeness to the population data. In a study (Takiar R 2023-1), an attempt was made to study the relationship between the Range and SD based on the 60 samples each of size 15, 30, 50, 75, 100, 125, 150 and 175 drawn from four different normal populations. The study revealed that the relationship between the SD and the Range can be characterized by the equation: Range = SD*(0.73*ln(n) + 1.72). In an another study, by examining the relationship between the Inter quartile Range (IQR) with that of Minimum and Maximum values seen in the samples (Takiar R 2023-2), a new method called the IQR-Takiar method is defined to identify the Outliers. The study revealed that the lower fence and higher fence values can be characterized as follows:

$$LF = Q_1 - IQR*[0.25*\ln(n)+0.20] \quad \text{and} \quad HF = Q_3 + IQR*[0.25*\ln(n)+0.20].$$

Based on the 20 Normal samples generated, the SD-Takiar method identified 29 Outliers while the IQR-Takiar method identified 27 Outliers and the IQR-Old method identified only 5 Outliers (Takiar R 2023-2). This raises a question whether the methods adopted by the Takiar identified correctly a greater number of Outliers or the IQR-Old method tend to pick up a smaller number of Outliers as compared to other two methods. In the absence of any clear cut evaluation method, it is difficult to decide which method should be adopted and which method should be adjudged as the appropriate one. Each method has its own theory and assumptions. The present study attempts to evolve an approach to demonstrate

empirically the superior method among the three to pick up Outliers from a set of data. The objectives of the present study therefore are:

- To develop an approach to identify the true Outliers.
- Use the SD-Range-Takiar method, IQR-Takiar method and IQR-Old method for detecting the Outliers from the selected sets of sampled data.
- Using the newly developed approach, adjudge the superior method among the three, for detecting the true Outliers.

Materials and Methods

The following three methods are employed to identify the possible Outliers in selected three normal samples.

IQR-Old Method: For this method, two fence values namely the Lower Fence value (LF) and the Higher Fence value (HF) are defined as follow:

$LF = Q_1 - 1.5*IQR$ and $HF = Q_3 + 1.5*IQR$

Where IQR is the Interquartile Range and Q_1 and Q_3 are the first and the third quartiles.

IQR-Takiar Method: Takiar defines the Lower Fence and Higher Fence values based on IQR as follows: $LF = Q_1 - IQR*[0.25*\ln(n)+0.20]$ and $HF = Q_3 + IQR*[0.25*\ln(n)+0.20]$ where n is the number of observations.

SD-Range-Takiar Method: The two fence values are defined as follows:

LF= Mean - SD*(0.37*ln (n) + 0.86) and HF = Mean + SD*(0.37*ln (n) + 0.86) where SD is derived with n as the denominator and not (n-1). In a study (Takiar R 2022), it was shown that SD with denominator (n-1) is not an unbiased estimate as claimed in theory and tend to overestimate the population variance by 17% consistently. In view of this, it is thought logical to use SD with n as the denominator.

Definition of Outliers: For Validation purposes, the artificially modified extreme values defined outside the original range are treated as true Outliers.

Selection of normal samples for detecting outliers

For the study purposes, three normal samples of size 15, 20 and 25 are generated using the function "Generation of Random Numbers" available on Excel. The samples were free from Outliers as evaluated by the selected three methods.

Table 1: Normal Samples according to their Size

Sample Size	Sample values														
15	42.1	47.3	49.6	50.6	52.8	54.2	55.6	58.4	58.8	59.3	59.6	68.4	68.9	74.2	74.6
20	20.0	23.5	23.6	23.7	29.2	33.0	36.5	38.6	38.8	40.0	46.2	46.8	47.7	48.4	48.7
	50.0	55.2	59.9	62.0	64.2										
25	31.5	32.7	36.2	47.5	50.2	53.3	56.8	58.4	68.2	69.7	69.8	71.2	72.2	75.1	75.9
	79.9	80.6	81.5	82.4	86.9	87.0	89.2	101.2	103.8	106.5					

Inrtoduction of outliers in the samples

In sample of size 15, the extreme values namely the minimum and maximum values are redefined by decreasing the minimum value and increasing the maximum value further. Proceeding in a similar way 10 new samples are obtained with only changed lower and higher values. The modified extreme values beyond the original sample range are treated as Outliers. Thus, two Outliers are introduced in each sample, each time. The aforementioned procedure was replicated for sample sizes of 20 and 25, resulting in the generation of ten new samples for each size. Three typical examples of redefined samples with introduction of Outliers, in the samples of size 15, 20 and 25 are shown in Table 2.

Table 2: Typical Examples of Redefined Samples with the Introduction of Outliers

Redefined samples	Sample values														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Original	**42.1**	47.3	49.6	50.6	52.8	54.2	55.6	58.4	58.8	59.3	59.6	68.4	68.9	74.2	**74.6**
1	**40.1**	47.3	49.6	50.6	52.8	54.2	55.6	58.4	58.8	59.3	59.6	68.4	68.9	74.2	**77.6**
2	**38.1**	47.3	49.6	50.6	52.8	54.2	55.6	58.4	58.8	59.3	59.6	68.4	68.9	74.2	**79.6**
3	**36.1**	47.3	49.6	50.6	52.8	54.2	55.6	58.4	58.8	59.3	59.6	68.4	68.9	74.2	**81.6**

Possible detections of outliers by the selected methods

Outliers by the IQR-Old method

IQR-Old method is employed to detect the introduced 20 outliers in 10 samples of size 15. The result of detection of Outliers by each sample is shown in Table 3. The method could detect 13 out of 20 Outliers suggesting 65% detection rate of Outliers. Even when the lower and higher values are changed by 6 numbers and the range is increased by 12 numbers, it went unnoticed by the method.

Table 3: Detection of Outliers by IQR-Old method – Sample size of 15

Variable		Sample with Changed Lower and Higher values									
	Original	1	2	3	4	5	6	7	8	9	10
Lower value	42.1	40.1	38.1	36.1	34.1	32.1	30.1	28.1	26.1	24.1	22.1
Higher Value	74.6	77.6	79.6	81.6	83.6	85.6	87.6	89.6	91.6	93.6	95.6
Outlier 1	-	-	-	-	-	32.1	30.1	28.1	26.1	24.1	22.1
Outlier 2	-	-	-	-	83.6	85.6	87.6	89.6	91.6	93.6	95.6

The Outliers detected in the sample size of 20 by the IQR-Old method are shown in Table 4. The method could detect 10 out of 20 Outliers suggesting the 50% detection rate of the Outliers. In this case, when the lower value is changed from 20.0 to 8.0 and the Higher value changed from 64.2 to 74, the method failed to pick up the changes in the data.

The Outliers detected in the sample size of 25 by the IQR-Old method is shown in Table 5. The method showed the 20% detection rate of the Outliers.

Table 4: Detection of Outliers by IQR-Old method – Sample size of 20

Variable	Sample value										
	Original	1	2	3	4	5	6	7	8	9	10
Lower value	20.0	15.0	14.0	15.0	10.0	8.0	6.0	5.0	4.0	2.0	0.0
Higher Value	64.2	65.2	68.0	70.0	72.0	74.0	76.0	77.0	78.0	80.0	82.0
Outlier 1	-	-	-	-	-	-	6.0	5.0	4.0	2.0	0.0
Outlier 2	-	-	-	-	-	-	76.0	77.0	78.0	80.0	82.0

Table 5: Detection of Outliers by IQR-Old method – Sample size of 25

Variable	Sample value										
	Original	1	2	3	4	5	6	7	8	9	10
Lower value	31.5	29.5	27.5	25.5	24.5	23.5	22.5	21.5	20.5	18.5	16.5
Higher Value	106.5	108.5	110.5	112.5	113.5	114.5	115.5	116.5	117.5	119.5	121.5
Outlier 1	-	-	-	-	-	-	-	-	-	-	16.5
Outlier 2	-	-	-	-	-	-	-	-	-	-	121.5

According to Table 5, when the lower value changed from 31.5 to 18.5 and the higher value changed from 106.5 to 119.5, the method could not detect the changes in the data showing it's insensitiveness to changes in extreme values and in detection of the Outliers.

Outliers by The IQR-Takiar Method

IQR-Takiar method showed 100% detection rate of the Outliers (Table 6).

Table 6: Detection of Outliers by IQR-Takiar method – Sample size of 15

Indicator	Sample value										
	Original	1	2	3	4	5	6	7	8	9	10
Lower value	42.1	40.1	38.1	36.1	34.1	32.1	30.1	28.1	26.1	24.1	22.1
Higher Value	74.6	77.6	79.6	81.6	83.6	85.6	87.6	89.6	91.6	93.6	95.6
Outlier 1	-	40.1	38.1	36.1	34.1	32.1	30.1	28.1	26.1	24.1	22.1
Outlier 2	-	77.6	79.6	81.6	83.6	85.6	87.6	89.6	91.6	93.6	95.6

IQR-Takiar method showed the 100% detection rate of the Outliers (Table 7).

Table 7: Detection of Outliers by IQR-Takiar method – Sample size of 20

Indicator	Sample value										
	Original	1	2	3	4	5	6	7	8	9	10
Lower value	20.0	15.0	14.0	15.0	10.0	8.0	6.0	5.0	4.0	2.0	0.0
Higher Value	64.2	65.2	68.0	70.0	72.0	74.0	76.0	77.0	78.0	80.0	82.0
Outlier 1	-	15.0	14.0	15.0	10.0	8.0	6.0	5.0	4.0	2.0	0.0
Outlier 2	-	65.2	68.0	70.0	72.0	74.0	76.0	77.0	78.0	80.0	82.0

IQR-Takiar method showed 100% detection rate of the Outliers (Table 8).

Table 8: Detection of Outliers by IQR-Takiar method – Sample size of 25

Indicator	Sample value										
	Original	1	2	3	4	5	6	7	8	9	10
Lower value	31.5	29.5	27.5	25.5	24.5	23.5	22.5	21.5	20.5	18.5	16.5
Higher Value	106.5	108.5	110.5	112.5	113.5	114.5	115.5	116.5	117.5	119.5	121.5
Outlier 1	-	29.5	27.5	25.5	24.5	23.5	22.5	21.5	20.5	18.5	16.5
Outlier 2	-	108.5	110.5	112.5	113.5	114.5	115.5	116.5	117.5	119.5	121.5

Outliers by The SD-Range-Takiar Method

SD-Range -Takiar method showed 95% detection rate (Table 9).

Table 9: Detection of Outliers by SD-Range-Takiar method – Sample size of 15

Indicator	Sample value										
	Original	1	2	3	4	5	6	7	8	9	10
Lower value	42.1	40.1	38.1	36.1	34.1	32.1	30.1	28.1	26.1	24.1	22.1
Higher Value	74.6	77.6	79.6	81.6	83.6	85.6	87.6	89.6	91.6	93.6	95.6
Outlier 1	-	-	38.1	36.1	34.1	32.1	30.1	28.1	26.1	24.1	22.1
Outlier 2	-	77.6	79.6	81.6	83.6	85.6	87.6	89.6	91.6	93.6	95.6

SD-Range -Takiar method showed 85% detection rate (Table 10).

Table 10: Detection of Outliers by SD-Range-Takiar method – Sample size of 20

Indicator	Sample value										
	Original	1	2	3	4	5	6	7	8	9	10
Lower value	20	15	14	12	10	8	8	5	4	2	0
Higher Value	64.2	65.2	68	70	72	74	72	77	78	80	82
Outlier 1	-	-	14	12	10	8	8	5	4	2	0
Outlier 2	-	-	-	70	72	74	72	77	78	80	82

SD-Range -Takiar method showed 65% detection rate (Table 11).

Table 11: Detection of Outliers by SD-Range-Takiar method – Sample size of 25

Indicator	Sample value										
	Original	1	2	3	4	5	6	7	8	9	10
Lower value	31.5	29.5	27.5	25.5	24.5	23.5	22.5	21.5	20.5	18.5	16.5
Higher Value	106.5	108.5	110.5	112.5	113.5	114.5	115.5	116.5	117.5	119.5	121.5
Outlier 1	-	-	27.5	25.5	24.5	23.5	22.5	21.5	20.5	18.5	16.5
Outlier 2	-	-	-	-	-	-	-	116.5	117.5	119.5	121.5

The Percentage detection of Outliers by the selected methods are shown in Fig. 1. The IQR-Old and SD-Range-Takiar method showed variations in detection of the Outliers by the sample size. Overall, when pooled for all the three sample sizes, the IQR-Old method could detect 42% of the 20 introduced Outliers while the SD-Range-Takiar method could detect 82% of the Outliers. The IQR-Takiar method showed 100% detection rate of Outliers. Hence, the results of the study suggests that IQR-Takiar method is a superior method among the three for detecting the Outliers.

Discussion

In literature there are few methods which are available to identify the Outliers from a given sample. While each method identify its own way the Outliers, very rarely they agree in total numbers of Outliers. Based on the 20 Normal samples generated, the SD-Takiar method identified 29 Outliers while the IQR-Takiar method identified 27 Outliers and the IQR-Old method identified only 5 outliers (Takiar R 2023-1, 2023-2). Based on the above observations,

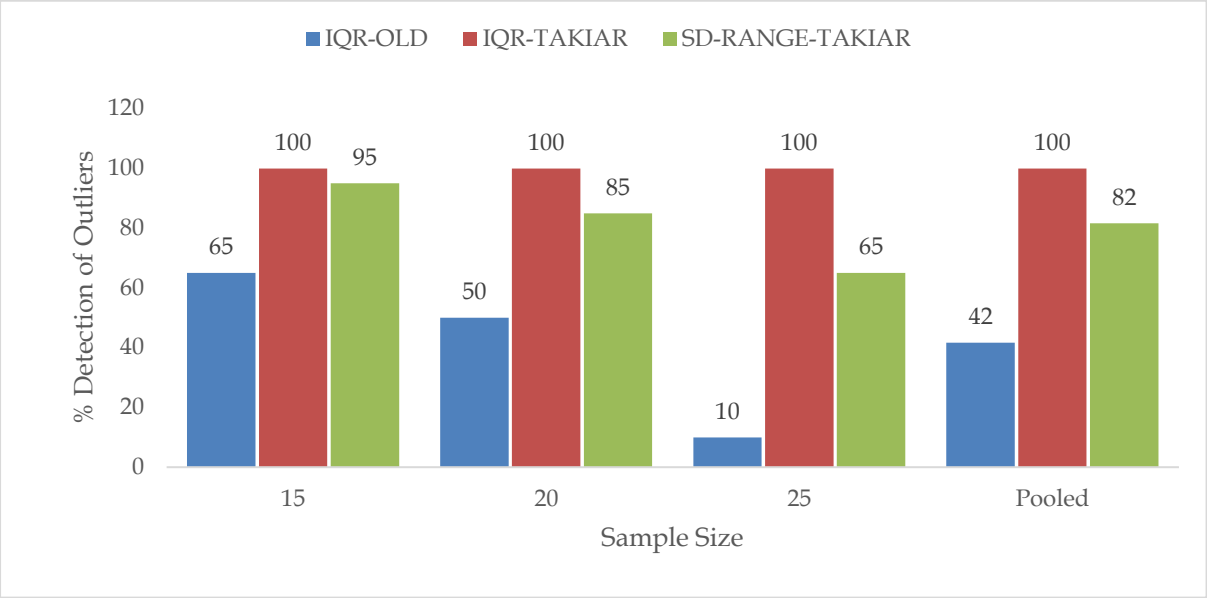

Figure 1: Percentage Detection of Outliers by the Methods

it can be argued that IQR-Old method identifies a smaller number of Outliers while others may say that SD-Range Takiar method is identifying a greater number of Outliers. Each method has its own assumptions and theory to support their results. But when on the same sample three different method identifies different number of Outliers, it is difficult to believe that all are correct. A minor variation in their number of Outliers will be acceptable to anyone but if the Outliers detected are 29, 27 and 5, it is difficult to decide which method should be adopted. The current study has adopted a novel approach to identify the true Outliers and also adjudged the superior method among the three for the identification of the Outliers.

The study started by choosing three random, normal samples of size 15,20 and 25 such that none of the IQR-Old, IQR-Takiar and SD-Range-Takiar method identifies any Outlier among them. It is logical to think that if the lower and higher values in the samples are replaced by the still lower and higher values, respectively then there should be no objection in treating them as the Outliers. Accordingly, lower and higher values in each sample are replaced by the still lower and the higher values, respectively and tested for identification of those Outliers. The IQR-Takiar method has shown 100% detection rate of Outliers while IQR-Old method showed the detection rate of 45%. This also suggests that IQR-Old method probably assign longer LF and HF interval than required. In view of the results of the study, it is advocated that IQR-Takiar method can safely be used to identify correctly the true Outliers in samples instead of the IQR-Old method.

Summary of The Observations

- In the present study, three method are used for the identification of Outliers from the selected three normal samples of size 15, 20 and 25.
- The three methods used are: IQR-Old method, IQR-Takiar method, SD-Range-Takiar method.

- For IQR-Old method, the Lower Fence (LF) and Higher Fence (HF) values are defined as follows: LF = Q_1 – 1.5*IQR and HF= Q_3+ 1.5* IQR.
- For IQR-Takiar method, the Fence values are defined as follows:

  LF = Q_1 – IQR*[0.25*ln(n)+0.20] and HF= Q_3+ IQR*[0.25*ln(n)+0.20]
- For SD-Range-Takiar method, the Fence values are defined as follows:

  LF= Mean - SD*(0.37*ln (n) + 0.86) and HF = Mean + SD*(0.37*ln (n) + 0.86)
- All the methods showed absence of Outliers in the initially selected samples of size 15, 20 and 25. .
- For each initial Outliers free sample, the minimum and maximum values are replaced by still lower and higher values thereby defining the new 10 samples with the Outliers.
- For Validation purposes, the artificially modified extreme values defined outside the original range are treated as the true Outliers.
- The Percentage detection rate of Outliers by the IQR-Old, SD-Range and IQR-Takiar method is observed to be 42%, 82% and 100%, respectively.
- The IQR-Takiar method is adjudged to be the superior method as compared to other two methods in detection of the Outliers.
- There is a need to evaluate the selected three methods for the identification of the Outliers for large samples, more than 30.
- The efficacy of this method remains to be evaluated when applied to non-normal populations, such as exponential and uniform distributions.

Recommendation

- The IQR-Takiar method is recommended to be used for the identification of Outliers in the samples.

References

[1]. Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics, 11*(1), 1–21. https://webspace.ship.edu/pgmarr/Geo441/Readings/Grubbs%201969%20-%20Detecting%20outlying%20observations%20in%20samples.pdf

[2]. Schwertman, N. C., Owens, M. A., & Adnan, R. (2004). A simple more general boxplot method for identifying outliers. *Computational Statistics & Data Analysis, 47*(1), 165–174. https://www.sciencedirect.com/science/article/abs/pii/S0167947303002469

[3]. Penn State Eberly College of Science. (2026). Identifying outliers: IQR method. *STAT 200: Elementary Statistics*. https://online.stat.psu.edu/stat200/lesson/3/3.2

[4]. Takiar, R. (2022). Sample variance – Is it really an unbiased estimate of the population variance? *Bulletin of Mathematics and Statistics Research, 10*(1), 21–30.

[5]. Takiar, R. (2023a). The relationship between the SD and the range and a method for the identification of the outliers. *Bulletin of Mathematics and Statistics Research, 11*(4), 62–75.

[6]. Takiar, R. (2023b). A new method to identify the outliers based on the inter quartile method. *Bulletin of Mathematics and Statistics Research, 11*(4), 103–114.

[7]. Vinutha, H. P., Poornima, B., & Sagar, B. M. (2018). Detection of outliers using interquartile range technique from intrusion dataset. In S. Satapathy, J. Tavares, V. Bhateja, & J. Mohanty (Eds.), *Information and decision sciences* (Advances in Intelligent Systems and Computing, Vol. 701). Springer. https://doi.org/10.1007/978-981-10-7563-6_53

Biography

Dr. Ramnath Takiar

I am a Post graduate in Statistics from Osmania University, Hyderabad. I did my Ph.D. from Jai Narain Vyas University of Jodhpur, Jodhpur, while in service, as an external candidate. I worked as a research scientist (Statistician) for Indian Council of Medical Research from 1978 to 2013 and retired from the service as Scientist G (Director Grade Scientist). I am quite experienced in large scale data handling, data analysis and report writing. I have 78 research publications , with 1376 citations to my credit, published in national and International Journals related to various fields like Nutrition, Occupational Health, Fertility and Cancer epidemiology. During the tenure of my service, I attended three International conferences namely in Goiana (Brazil-2006), Sydney (Australia-2008) and Yokohoma (Japan-2010) and presented a paper in each. I also attended the Summer School related to Cancer Epidemiology (Modul I and Module II) conducted by International Agency for Research in Cancer (IARC), Lyon, France from 19th to 30th June 2007. After my retirement, I joined my son at Ulaanbaatar, Mongolia. I worked in Ulaanbaatar as a Professor and Consultant from 2013-2018 and was responsible for teaching and guiding the Ph.D. students. I also taught Mathematics to undergraduates and Econometrics to MBA students. During my service there, I also acted as the Executive Editor for the in-house Journal "International Journal of Management". I am also acting as a reviewer for few International Journals. I am still active in research and have published 17 research papers in Statistical Methodologies during 2021-25.