

---

# Outlier Detection Technique for Univariate Normal Datasets

Ooko Silas Owuor<sup>1</sup>, Troon John Benedict<sup>2</sup>, Otieno Okumu Kevin<sup>2</sup>

<sup>1</sup>Department of Mathematics and Physical Sciences, Maasai Mara University, Narok, Kenya

<sup>2</sup>Department of Economics, Maasai Mara University, Narok, Kenya

## Email address:

sylusooko7@gmail.com (O. S. Owuor), troon@mmarau.ac.ke (T. J. Benedict), kevinotieno15@gmail.com (O. O. Kevin)

## To cite this article:

Ooko Silas Owuor, Troon John Benedict, Otieno Okumu Kevin. Outlier Detection Technique for Univariate Normal Datasets. *American Journal of Theoretical and Applied Statistics*. Vol. 11, No. 1, 2022, pp. 1-12. doi: 10.11648/j.ajtas.20221101.11

**Received:** December 19, 2021; **Accepted:** January 8, 2022; **Published:** January 21, 2022

---

**Abstract:** This paper presents an outlier detection technique for univariate normal datasets. Outliers are observations that lie at an abnormal distance from the mean. Outlier detection is a useful technique in such areas as fraud detection, financial analysis, health monitoring and Statistical modelling. Many recent approaches detect outliers according to reasonable, pre-defined concepts of an outlier. Methods of outlier detection such as Gaussian method of outlier detection have been widely used in the detection of outliers for univariate data-sets, however, such methods use measure of central tendency and dispersion that are affected by outliers hence making the method to be less robust towards detection of outliers. The study aimed at providing an alternative method that can be used in outlier detection for univariate normal data sets by deploying the measures of variation and central tendency that are least affected by the outliers (median and the geometric measure of variation). The study formulated an outlier detection formula using median and geometric measure of variation and then applied the formulation on randomly simulated normal dataset with outliers and recorded the number of outliers detected by the method in comparison to the other two existing best methods of outlier detection. The study then compared the sensitivity of the three methods in outlier detection. The simulation was done in two different ways, the first considered the variation in mean with a constant standard deviation while the second test held the mean constant while varying the standard deviation. The formulated outlier detection technique performed the best, eliminating the most required number of outliers compared to other two Gaussian outlier detection techniques when there was variation in mean. The study also established that the formulated method of outlier detection was stricter when the standard deviation was varied but still stands out to be the best as an outlier is defined relative to the mean and not the standard deviation. The study established that the formulated method is more sensitive than the Gaussian Method of outlier detection but performed as well as the best existing outlier detection technique. In conclusion, the study established that the formulated method could be employed in outlier detections for univariate normal data-sets as it performed almost the same to the best existing method of outlier detection for univariate data-sets.

**Keywords:** Outlier, Anomaly, Outlier Detection, Gaussian

---

## 1. Introduction

Outlier detection; also known as anomaly detection this process is the identification of rare items [1, 2] events and observations which arise and are significantly different from the other observations in the data [3, 4]. Identification of these events (outliers) is very important given they may lead to bad data and this may lead to poor running of the experiment for they may hide very essential information about the data. If it can be determined earlier that a point is outlying then it can be worth ejecting it for the purposes of better results. Secondly, in some cases, it may not be possible to deter-

mine if an outlying point is bad data since Outliers may be due to random variation or may indicate something scientifically interesting. In any event, we typically do not want to simply delete the outlying observation. However, if the data contains significant outliers, we may need to consider the use of robust statistical techniques. [5-7] Before application of these techniques we have to determine whether the outlier is univariate or multivariate. Univariate outliers can be found when looking at a distribution of values in a single feature space. Multivariate outliers can be found in an n-dimensional space (of n-features). Looking at distributions in n-dimensional spaces can be very difficult for the human brain, that is why we need to train a model to do it for us [8,

9]. Outlier detection is an important research problem in data mining that aims to find objects that are considerably dissimilar, exceptional and inconsistent with respect to the majority data in an input database [10]. The following are the existing outlier detection techniques that the study focused on.

**1.1. Gaussian Model**

Estimation of mean and standard deviation is done in training stage using the maximum likelihood estimates (MLE). A wide range, nearly 100 of outlier tests has been put in place in different ways depending on the data set and the parameters like mean and variance and the expected values of the outliers. To ensure the test carried are optima or close to optima statistical discordancy tests are usually carried out in the test stage [11-13]. The usually used outlier test for normal distribution is the mean-variance and Boxplot tests. In the mean variance test for Gaussian distribution  $N(\mu, \sigma)$ , where the population has mean and variance  $\sigma$ . Outlier is considered to be a point that lie 3 or more standard deviation i.e.  $> 3\sigma$  away from the mean.

Similarly, the tests can be applied to some other distribution like t-distribution and the Poisson distribution with the former featuring a latter tail and the latter a longer right tail than a normal distribution.

The box plot test also gives a profound test by deployment of 5 major attributes i.e. smallest non-outlier observation [min], lower quartile [Q1], upper quartile [Q3], medium and the largest non-outlier observation [max]. The quantity (Q3 - Q1) is called the interquartile range (IQR). This helps use clearly define the boundary beyond which the data will be considered an outlier. A point  $X_1$  is labeled or referred to as an outlier if,  $X_i > Q_3 + k(IQR)$  or  $X_i < Q_1 - k(IQR)$ .

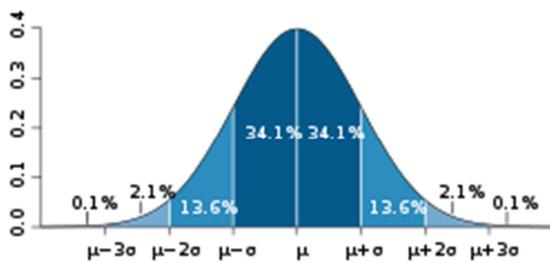


Figure 1. Outliers are points  $> |\mu + 3\sigma|$ , for some  $k=1.5$ .

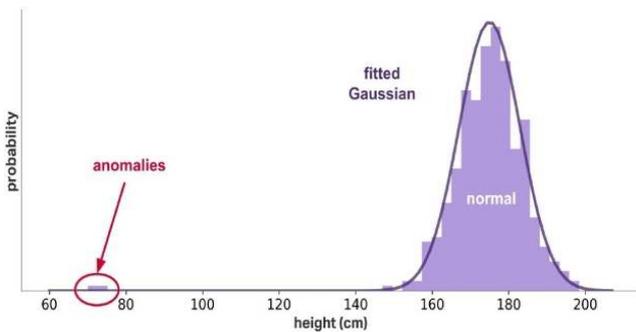


Figure 2. Outlier detection by fitted Gaussian model.

Basing our argument on low dimensional outlier detection technique, we settle with the Gaussian model which defines an outlier as a point  $X > |\mu + 3\sigma|$  as the best existing detection technique as it takes into consideration both the probabilistic and normal distributions.

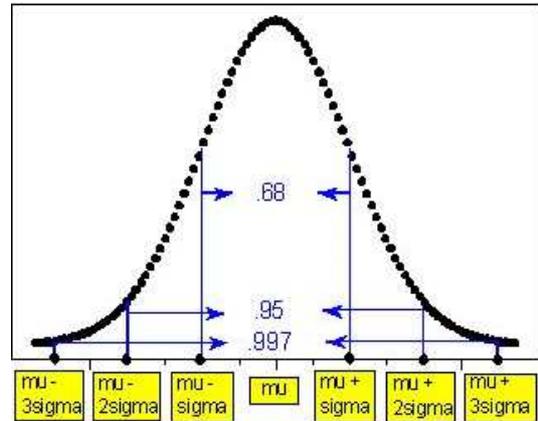


Figure 3. Outliers are points  $> |\mu + 3\sigma|$ .

However, this method has a number of shortcomings since by Central Limit Theorem (Which states; If you have small, independent random variables, then their sum is distributed approximately a bell curve [14-16]). By so doing, if an outlier occurs at some point away from the normal curve, then the normal curve will shift towards the outlier.

The anomaly/outlier towards the left as shown in Figure 5 will shift the normal curve towards the left as illustrated below;

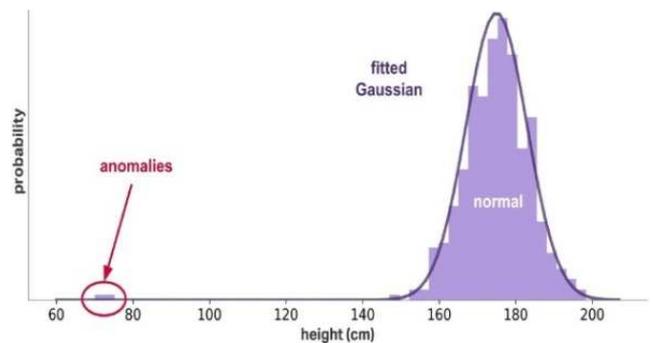


Figure 4. Positioning of an outlier towards the left of the curve.

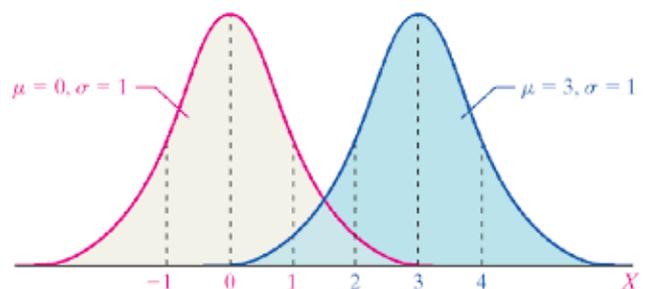


Figure 5. Shift of the normal curve towards the outlier changing the mean but keeping the standard deviation constant.

In an event outliers occur both side of the curve then it's likely to spread the normal curve having an effect on the

standard deviation but keeping the mean constant. This is illustrated in the Figure 7 below;

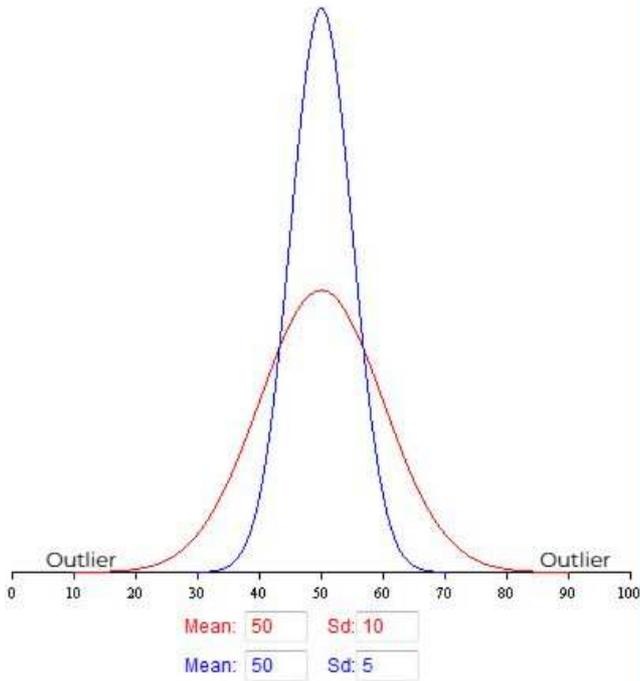


Figure 6. Stretching of the curve when outliers occur at both sides of the curve.

Since the existing Gaussian model detection technique uses parameters that are affected by the outliers as illustrated above, the study would like to come up with a technique that does not rely on these parameters  $\mu$  and  $\sigma$ . In this regard, the study will replace the former with Median since the Median is least affected by the outliers [17, 18] and the latter with the Geometric measure because it takes into account the compounding that occurs from period to period. Because of this, investors usually consider the geometric mean a more accurate measure of returns than the arithmetic mean [19].

In so doing we will have our new detection technique define outlier as a point  $X > |Med + 3g|$

Where;

Med is the median and g is the geometric measure. The main objectives of this study is Outlier detection for univariate data set using geometric technique.

### 1.2. Regression Model

A regression model is also used to detect the outliers. In this scenario, an outlier is considered to be an observation for which the residual is larger compared to other observations in the data-set. Such observations are imputed accordingly for higher accuracy in statistical findings. This study however is going to focus on the Gaussian detection techniques.

## 2. Methods

For normal observations, the outlier detection technique by Gaussian model stipulates that an outlier is given by a

point  $X > |\bar{X} \pm 3\sigma|$ . Since the arithmetic mean as a measure of central tendency that is affected by the outliers, and we know that in a perfectly symmetric data, the mean, the mode and the median are the same [20, 21], the study replaced the mean with the median since the median is a measure of central tendency that is not affected by the existence of the outliers in the dataset [22]. This lead to the same equation given as;

$$X > |Med \pm 3\sigma| \tag{1}$$

The study expects the formula to be better than the Gaussian outlier detection method given we have done away with a measure of central tendency that is affected by the outliers. Moreover, since the standard deviation is a measure of variation that is affected by the existence of the outliers in the dataset and definitely will affect the accuracy of the detection, the study therefore found it necessary to replace the standard deviation with a geometric mean. The geometric mean however was calculated around the median, a measure of central tendency that is not affected by the outliers [22, 23], so as to come up with a geometric mean that is also not affected by the existence of the outliers in the dataset. The study expects this to make our outlier detection formula even better. The formula will therefore be given as;

$$X > |Med \pm 3G| \tag{2}$$

The next task now is to calculate the geometric averages with respect to the mean, this is given as, the study borrowed a concept from [24]

$$G = \sqrt[n]{\prod_{i=1}^n (xi - Med)}$$

While formulating the G, the study established that most important is the deviation from the median can either take a positive, a negative or a zero value, making the formula not applicable in an event we get a negative value since we cannot get a real root of a negative number. In response to this shortcoming, the study took the absolute of the deviations given the rule of geometric averaging holds that most important is the magnitude of the deviation and not the direction [25, 26]. By doing so, we obtain the equation as;

$$G = \sqrt[n]{\prod_{i=1}^n |xi - Med|}$$

The next challenge comes when some data points are equal to the median leading to zero deviation, thus, the product of the deviation from the median will eventually be zero leading to indefinite root. In response to this problem, the study added an arbitrary constant k. The study derived constant K by identifying the best constant that best detects outliers in low dimensional data-set and as well have the least effect to the deviation, the constant was obtained by plotting these constants against outliers removed in given sets of data. This is as shown in Figure 7 below;

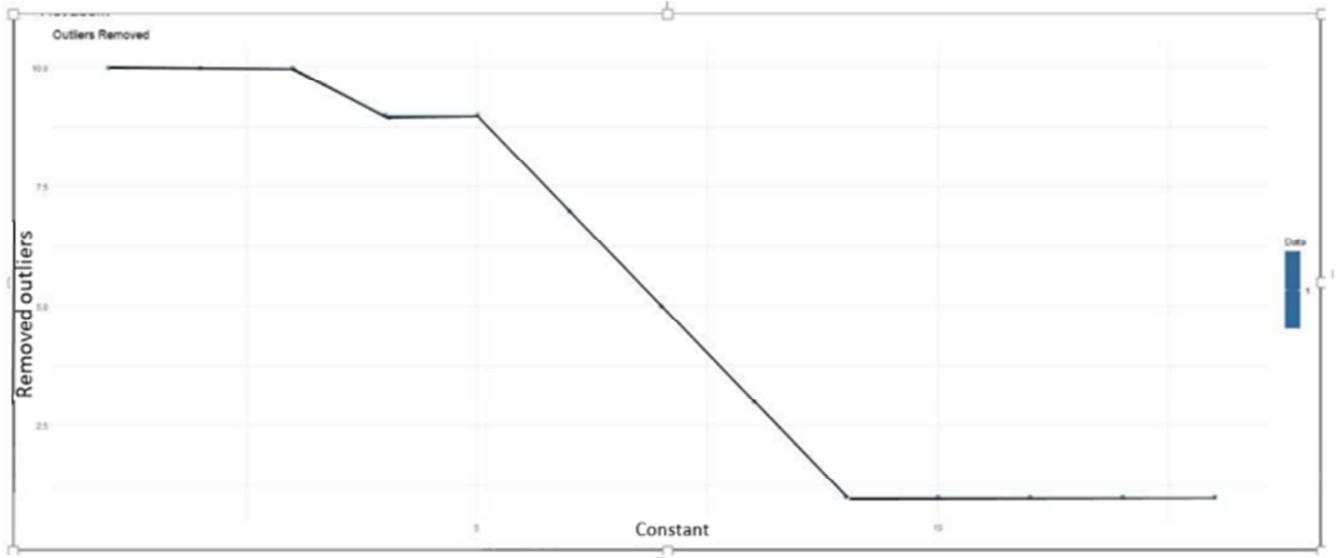


Figure 7. Curve used to derive the most appropriate k-value.

The curve flattened at the 9<sup>th</sup> constant which was 0.1, even though the other constants removed the same number of outliers after flattening, the study considered the least of these constants (0.1) which will have the least effect on the deviation from the median given the study doesn't want to interfere much with the deviation from the median. The formula therefore becomes;

$$G = \sqrt[n]{\prod_{i=1}^n \frac{|di| + k, i = 1 \dots n}{xi = Med}}$$

Where  $di=|xi-Med|$  and  $k=0.1$

The study further introduced logarithms in order to help in eliminating infinite number that are likely to arise when the population size is large as the geometric measure of variation from the median is being calculated. This gives;

$$G = \begin{cases} \exp\left(\frac{\sum_{i=1}^n \log(|xi-Med|+k)}{n}\right), & x1 \neq x2 \neq \dots \neq Med \\ 0, & x1 = x2 = \dots = Med \end{cases}$$

Where  $k=0.1$

Therefore the new formula for outlier detection will state that an outlier is any point given as;

$$X > |Med \pm 3G|$$

### 3. Results

To test for the effectiveness of the newly invented formula, the study examined and compared the sensitivity of the two Gaussian detection techniques ( $X_i > Q3 + k(IQR)$  or  $X_i < Q1 - k(IQR)$ ).

The study formulated random normal data-sets to help in randomly obtaining the data for simulation. This was done by combining two randomly formulated data-sets with varying mean and standard deviation to help examine the effect of changing either the mean or the standard deviation to the sensitivity of the model. Given the Gaussian formula under scrutiny ( $X > |\mu + 3\sigma|$ ) uses two measure of central tendency that are affected by the outliers, this process was done in two ways;

1. Combining two data-sets with constant standard deviation but varying mean

The first simulation is summarized in the table below;

Table 1. Outliers detected by all formulas with first set of data.

Technique	New formula	Gaussian 1st equation	Gaussian 2 <sup>nd</sup> Equation
Outliers available	5	5	5
Outliers removed	7	6	4

From table 1, the first simulation, the study combined the first data set ( $X1; n=50, \mu = 4$  and  $\sigma = 2$ ) with ( $X2; n=5, \mu = 30$  and  $\sigma = 2$ ), with ( $X2; n=5, \mu = 30$  and  $\sigma = 2$ ). The expected 5 outliers from the combined datasets were then subjected to the three detection techniques. The study simulated the dataset in the first Gaussian detection technique ( $X_i > Q3 + k(IQR)$ ) or  $X_i < Q1 - k(IQR)$ ), 6 outliers were detected as shown in the table above (Refer to Appendix 1: Figure 8). When the same dataset was simulated against the

second Gaussian equation ( $X > |\mu + 3\sigma|$ ), 4 outliers were detected (Refer to Appendix 1: Figure 9).

When the outliers were simulated in the new equation, 7 outliers were detected (refer to Appendix 1: Figure 10).

From the results, the new formula eliminated the most number of outliers (7), the number exceeds the expected number of outliers, 5, since when two datasets of different means are combined they form a new mean, therefore the outliers are likely to be more or less than were suppose.

The Gaussian second equation performed second best with 6 outliers eliminated. The Gaussian equation under scrutiny managed to detect only 4 outliers, this may be attributed to by the fact that it use the measures of central tendency that is

affected by the outliers.

The study examined another pair of datasets, (X3; n=250,  $\mu = 15$  and  $\sigma = 5$ ) and (X4; n=10,  $\mu = 80$  and  $\sigma = 5$ ).

The table below summarizes the simulations.

*Table 2. Outliers detected by all formulas with second set of data.*

Technique	New formula	Gaussian 1st equation	Gaussian 2 <sup>nd</sup> Equation
Outliers available	10	10	10
Outliers removed	22	12	10

From table 2, when the data was simulated using the first Gaussian formula (refer to Appendix 2: Figure 11), 10 outliers were detected. When the data was simulated against the second Gaussian formula (refer to Appendix 2: Figure 12), 10 outliers were detected.

Finally, when the dataset was simulated against the new detection technique and 22 outliers were detected.

The new formula eliminated the more outliers compared to the rest which may be contributed to by the fact that it uses measures that are least affected by the outliers. The Gaussian equation with the interquartile range once again performed better, eliminating 12 outliers. The equation under scrutiny

(the second Gaussian equation) eliminated the least number of outliers.

The number of outliers removed differ and may be even more than expected because the moment two datasets are combined, they form a new mean interfering with the number of outliers as outliers by definition, are observations that lip an abnormal distance from the mean.

The study carried another sensitivity test on another sets of datasets by combining (X5; n=150,  $\mu = 90$  and  $\sigma = 5$ ) with (X6; n=25,  $\mu = 200$  and  $\sigma = 5$ ).

The summary is as shown in the table below;

*Table 3. Outliers detected by all formulas with third set of data.*

Technique	New formula	Gaussian 1st equation	Gaussian 2 <sup>nd</sup> Equation
Outliers available	10	10	10
Outliers removed	27	26	26

When the outliers were simulated using the first Gaussian equation ( $X_i > Q_3 + k(IQR)$  or  $X_i < Q_1 - k(IQR)$ ), 26 outliers were eliminated (Refer to Appendix 3: Figure 14).

When the outliers were simulated using the Gaussian second equation, 27 outliers (refer to Appendix 3: Figure 15).

Finally, when the data was simulated against the new detection technique, 26 outliers were detected (Refer to Appendix 3: Figure 16).

The three techniques performed relatively the same even though the new technique eliminated one more outlier than the rest.

*2. Combining two data-sets with constant mean but varying standard deviation*

The study examined the sensitivity by combining (X7; n=250,  $\mu = 40$  and  $\sigma = 45$ ) with (X8; n=15,  $\mu = 40$  and  $\sigma = 5$ ).

The finding are summarized in the table as shown;

*Table 4. Outliers detected by all formulas with first set of data.*

Technique	New formula	Gaussian 1st equation	Gaussian 2 <sup>nd</sup> Equation
Outliers available	15	15	15
Outliers removed	27	4	1

Simulating the outliers using the Gaussian first equation, 4 outliers were removed. The Gaussian second equation eliminated 1 outlier and the new detection technique removed 27 outliers.

The study also examined the following datasets;

(X9; n=500,  $\mu = 20$  and  $\sigma = 10$ ) and (X10; n=55,  $\mu = 20$  and  $\sigma = 5$ ). The summary is as shown in the table below;

*Table 5. Outliers detected by all formulas with second set of data..*

Technique	New formula	Gaussian 1st equation	Gaussian 2 <sup>nd</sup> Equation
Outliers available	55	55	55
Outliers removed	72	8	3

Simulating the outliers using the Gaussian first equation, 4 outliers were ejected from the data-set, the Gaussian second equation detected 1 outlier while the new equation once again removed the most, 27. This may be as a result the use of measures of central tendencies that are least affected by

the outliers. Lastly, the study examined the sensitivity in one more pair of data-sets; (X9; n=500,  $\mu = 20$  and  $\sigma = 10$ ) with. When the outliers were detected, the Gaussian first equation ejected 8 outliers, 3 by the second equation and 72 by the new equation. The new equation is stricter because the

measures it uses are not affected by the outliers.

## 4. Summary

The study sought to determine the Outlier detection for univariate data set using geometric technique. The study sought to empirically detect outliers using the univariate normal outlier detection technique in simulated data. The study also aim to measure precision of the univariate outlier detection model in comparison to the Gaussian outlier detection models. The data used in this study was randomly generated from a normal distribution. This chapter gives a summary of the findings, makes conclusions and recommendations based on the findings.

### 1) Summary of findings

The study sought to establish an outlier detection technique for univariate normal datasets. The measures of central tendency in the Gaussian equation ( $X > \mu + 3\sigma$ ) which are highly affected by the outliers were replaced by those that are least affected by the outliers to form a new equation which defined an outlier as a point  $X > Med + 3G$ .

The study sought to empirically detect outliers using univariate normal outlier detection technique in simulated data. The normal data sets were randomly generated and

simulation done using the new formula, the study noted that the formula was able to detect the outliers in the data-set. Univariate normal outlier detection model in comparison to the Gaussian outlier detection model. The study formulated same sets of datasets and observed the sensitivity of the models.

### 2) Conclusion

After conducting sensitivity test on several sets of datasets, the study established that the new formula is the best in outlier detection, in all cases examined it performed better than the Gaussian detection model ( $X > \mu + 3\sigma$ ). The second Gaussian equation ( $X_i > Q_3 + k(IQR)$  or  $X_i < Q_1 - k(IQR)$ ) performed as well as the new formula ( $X > Med \pm 3G$ ).

When there was variation in mean but constant standard deviation. This, however, was not the case when standard deviation was varied with constant mean, in this case the new model detected more outliers than any of the two Gaussian equations. This is due to the fact that the new equation used on the measures that are least affected by the outliers.

The new equation proved more sensitive and precise in outlier detection. Even though the new technique was stricter when the standard deviation was varied, it still stands as the best technique according to the study as an outlier is defined relative to the mean and not the standard deviation.

## Appendix

### Appendix 1. Outlier Detection Simulation for Small Data Sets with Mean Variation

```
> set.seed(45)
> x1<-rnorm(50,mean=4,sd=2)
> x2<-rnorm(5,mean=30,sd=2)
> v=c(x1,x2)
> v[v<quantile(v,.25)-1.5*IQR(v) | v>quantile(v,.75)+1.5*IQR(v)] <- NA #Gaussian Eqn 1
v
[1] 4.681599 2.593319 3.240925 2.507905 2.203785 3.330412 2.997244 3.650929 7.618075 3.539790 1.739164 4.431978 6.464475 7.218717
[15] 4.803101 3.454032 3.927695 3.699378 NA 0.695008 1.729710 4.455340 3.633363 3.172963 3.124809 3.947631 2.280332 4.333089
[29] 6.950981 4.390846 4.318844 2.559613 2.128995 4.570865 2.521530 4.858298 9.467968 1.333193 7.720191 4.491940 2.508021 1.031724
[43] 4.444097 4.955655 5.462397 4.344210 6.373384 3.299181 6.295195 6.700168 NA NA NA NA NA
> sum(is.na(v))
[1] 6
```

Figure 8. Sensitivity of Gaussian 1<sup>st</sup> equation with variation in mean.

```
> set.seed(45)
> x1<-rnorm(50,mean=4,sd=2)
> x2<-rnorm(5,mean=30,sd=2)
> v=c(x1,x2)
> v[v< mean(v)-3*sd(v) | v> mean(v)+3*sd(v)] <- NA #Gaussian Eqn 2
v
[1] 4.681599 2.593319 3.240925 2.507905 2.203785 3.330412 2.997244 3.650929 7.618075 3.539790 1.739164 4.431978
[13] 6.464475 7.218717 4.803101 3.454032 3.927695 3.699378 11.537621 0.695008 1.729710 4.455340 3.633363 3.172963
[25] 3.124809 3.947631 2.280332 4.333089 6.950981 4.390846 4.318844 2.559613 2.128995 4.570865 2.521530 4.858298
[37] 9.467968 1.333193 7.720191 4.491940 2.508021 1.031724 4.444097 4.955655 5.462397 4.344210 6.373384 3.299181
[49] 6.295195 6.700168 NA NA NA 26.944019 NA
> sum(is.na(v))
[1] 4
```

Figure 9. Sensitivity of Gaussian 2<sup>nd</sup> equation with variation in mean.

```
> set.seed(45)
> x1<-rnorm(50,mean=4,sd=2)
> x2<-rnorm(5,mean=30,sd=2)
> v=c(x1,x2)
> v[v<quantile(v,.50)-3*exp(sum(Log(abs(v-quantile(v,.50))+1))/length(v)) |
+ v>quantile(v,.50)+3*exp(sum(Log(abs(v-quantile(v,.50))+1))/length(v))] <- NA #New Eqn
v
[1] 4.681599 2.593319 3.240925 2.507905 2.203785 3.330412 2.997244 3.650929 7.618075 3.539790 1.739164 4.431978 6.464475 7.218717
[15] 4.803101 3.454032 3.927695 3.699378 NA 0.695008 1.729710 4.455340 3.633363 3.172963 3.124809 3.947631 2.280332 4.333089
[29] 6.950981 4.390846 4.318844 2.559613 2.128995 4.570865 2.521530 4.858298 NA 1.333193 7.720191 4.491940 2.508021 1.031724
[43] 4.444097 4.955655 5.462397 4.344210 6.373384 3.299181 6.295195 6.700168 NA NA NA NA NA
> sum(is.na(v))
[1] 7
```

Figure 10. Sensitivity of the New Equation with variation in mean.



### Appendix 3. Outlier Detection Simulation for Large Data Sets with Mean Variation

```

> set.seed(45)
> x7<-rnorm(150,mean=90,sd=5)
> x8<-rnorm(25,mean=200,sd=5)
> v=c(x7,x8)
> v[v<quantile(v,.25)-1.5*IQR(v) | v>quantile(v,.75)+1.5*IQR(v)] <- NA #Gaussian Eqn 1
> v
[1] 91.70400 86.48330 88.10231 86.26976 85.50946 88.32603 87.49311 89.12732 99.04519 88.84948 84.34791 91.07994
[13] 96.16119 98.04679 92.00775 88.63508 89.81924 89.24844 NA 81.73752 84.32427 91.13835 89.08341 87.93241
[25] 87.81202 89.86908 85.70083 90.83272 97.37745 90.97711 90.79711 86.39903 85.32249 91.42716 86.30382 92.14574
[37] 103.66992 83.33298 99.30048 91.22985 86.27005 82.57931 91.11024 92.38914 93.65599 90.86053 95.93346 88.24795
[49] 95.73799 96.75042 95.58075 91.01173 90.55745 82.36005 90.28754 100.93730 89.64523 96.46498 91.88438 85.91745
[61] 89.64153 79.20747 100.09597 89.74935 88.59893 91.58919 85.15548 87.49829 85.19538 82.04977 91.87867 95.14599
[73] 96.13116 95.24197 90.83707 95.68999 95.28363 86.59921 83.31305 80.98449 93.50489 79.77955 85.05972 93.13691
[85] 88.08129 108.01761 91.66105 83.77471 91.23683 82.97986 94.53591 86.48023 82.15114 91.72656 93.39969 95.71359
[97] 92.02958 85.40680 101.48157 86.55151 80.60851 95.32172 95.69164 93.89804 86.83979 90.94214 85.45155 89.31991
[109] 91.14678 86.37434 93.80565 89.60261 98.63746 92.02601 82.98189 90.33236 91.52715 94.36589 89.85526 90.65314
[121] 90.37835 93.93083 81.86099 90.39073 86.92468 90.46637 90.82302 84.63411 88.12258 89.90127 93.04207 92.37104
[133] 81.54538 94.60573 92.60130 92.17799 88.80717 84.58242 93.90847 86.33207 90.91520 92.11036 89.86157 92.27670
[145] 96.28178 83.34470 85.56218 89.47476 89.54752 80.99463 NA NA NA NA NA NA
[157] NA NA
[169] NA NA
> sum(is.na(v))
[1] 26

```

Figure 14. Sensitivity of Gaussian 1<sup>st</sup> equation with variation in mean.

```

> set.seed(45)
> x7<-rnorm(150,mean=90,sd=5)
> x8<-rnorm(25,mean=200,sd=5)
> v=c(x7,x8)
> v[v< mean(v)-3*sd(v) | v> mean(v)+3*sd(v)] <- NA #Gaussian Eqn 2
> v
[1] 91.70400 86.48330 88.10231 86.26976 85.50946 88.32603 87.49311 89.12732 99.04519 88.84948 84.34791 91.07994
[13] 96.16119 98.04679 92.00775 88.63508 89.81924 89.24844 108.84405 81.73752 84.32427 91.13835 89.08341 87.93241
[25] 87.81202 89.86908 85.70083 90.83272 97.37745 90.97711 90.79711 86.39903 85.32249 91.42716 86.30382 92.14574
[37] 103.66992 83.33298 99.30048 91.22985 86.27005 82.57931 91.11024 92.38914 93.65599 90.86053 95.93346 88.24795
[49] 95.73799 96.75042 95.58075 91.01173 90.55745 82.36005 90.28754 100.93730 89.64523 96.46498 91.88438 85.91745
[61] 89.64153 79.20747 100.09597 89.74935 88.59893 91.58919 85.15548 87.49829 85.19538 82.04977 91.87867 95.14599
[73] 96.13116 95.24197 90.83707 95.68999 95.28363 86.59921 83.31305 80.98449 93.50489 79.77955 85.05972 93.13691
[85] 88.08129 108.01761 91.66105 83.77471 91.23683 82.97986 94.53591 86.48023 82.15114 91.72656 93.39969 95.71359
[97] 92.02958 85.40680 101.48157 86.55151 80.60851 95.32172 95.69164 93.89804 86.83979 90.94214 85.45155 89.31991
[109] 91.14678 86.37434 93.80565 89.60261 98.63746 92.02601 82.98189 90.33236 91.52715 94.36589 89.85526 90.65314
[121] 90.37835 93.93083 81.86099 90.39073 86.92468 90.46637 90.82302 84.63411 88.12258 89.90127 93.04207 92.37104
[133] 81.54538 94.60573 92.60130 92.17799 88.80717 84.58242 93.90847 86.33207 90.91520 92.11036 89.86157 92.27670
[145] 96.28178 83.34470 85.56218 89.47476 89.54752 80.99463 203.55123 200.20106 201.49178 206.21652 198.23154 185.74605
[157] 204.39366 189.76887 208.22311 197.90699 204.58194 201.80422 211.02346 204.32687 207.48455 203.35738 193.95942 197.29920
[169] 195.43880 194.67938 194.23424 201.54845 200.59899 203.08030 195.52335
> sum(is.na(v))
[1] 0

```

Figure 15. Sensitivity of Gaussian 2<sup>nd</sup> equation with variation in mean.

```

> set.seed(45)
> x7<-rnorm(150,mean=90,sd=5)
> x8<-rnorm(25,mean=200,sd=5)
> v=c(x7,x8)
> v[v<quantile(v,.50)-3*exp(sum(log(abs(v-quantile(v,.50))+.1))/length(v)) |
+ v>quantile(v,.50)+3*exp(sum(log(abs(v-quantile(v,.50))+.1))/length(v))] <- NA #New Eqn
> v
[1] 91.70400 86.48330 88.10231 86.26976 85.50946 88.32603 87.49311 89.12732 99.04519 88.84948 84.34791 91.07994
[13] 96.16119 98.04679 92.00775 88.63508 89.81924 89.24844 NA 81.73752 84.32427 91.13835 89.08341 87.93241
[25] 87.81202 89.86908 85.70083 90.83272 97.37745 90.97711 90.79711 86.39903 85.32249 91.42716 86.30382 92.14574
[37] 103.66992 83.33298 99.30048 91.22985 86.27005 82.57931 91.11024 92.38914 93.65599 90.86053 95.93346 88.24795
[49] 95.73799 96.75042 95.58075 91.01173 90.55745 82.36005 90.28754 100.93730 89.64523 96.46498 91.88438 85.91745
[61] 89.64153 79.20747 100.09597 89.74935 88.59893 91.58919 85.15548 87.49829 85.19538 82.04977 91.87867 95.14599
[73] 96.13116 95.24197 90.83707 95.68999 95.28363 86.59921 83.31305 80.98449 93.50489 79.77955 85.05972 93.13691
[85] 88.08129 NA 91.66105 83.77471 91.23683 82.97986 94.53591 86.48023 82.15114 91.72656 93.39969 95.71359
[97] 92.02958 85.40680 101.48157 86.55151 80.60851 95.32172 95.69164 93.89804 86.83979 90.94214 85.45155 89.31991
[109] 91.14678 86.37434 93.80565 89.60261 98.63746 92.02601 82.98189 90.33236 91.52715 94.36589 89.85526 90.65314
[121] 90.37835 93.93083 81.86099 90.39073 86.92468 90.46637 90.82302 84.63411 88.12258 89.90127 93.04207 92.37104
[133] 81.54538 94.60573 92.60130 92.17799 88.80717 84.58242 93.90847 86.33207 90.91520 92.11036 89.86157 92.27670
[145] 96.28178 83.34470 85.56218 89.47476 89.54752 80.99463 NA NA NA NA NA NA
[157] NA NA
[169] NA NA
> sum(is.na(v))
[1] 27

```

Figure 16. Sensitivity of the New Equation with variation in mean.

Appendix 4. Outlier Detection Simulation for Large Data Sets with Standard Deviation Variation

```

> set.seed(45)
> x9<-rnorm(230,mean=40,sd=45)
> x10<-rnorm(15,mean=40,sd=5)
> v=c(x9,x10)
> v[v<quantile(v,.25)-1.5*IQR(v) | v>quantile(v,.75)+1.5*IQR(v)] <- NA #Gaussian Eqn 1
> v
[1] 55.33598609 8.34968643 22.92080212 6.42786532 -0.41482985 24.93426513 17.43798310 32.14589365 121.40668314
[10] 29.64527615 -10.86881983 49.71949991 95.45067783 112.42114181 58.06977820 27.71571846 38.37314459 33.23599445
[19] NA -34.36231895 -11.08152960 50.24515759 31.75066583 21.39166189 20.30821228 38.82170406 1.30746202
[28] 47.49450598 106.39708285 48.79403115 47.17398058 7.59130220 -2.09761441 52.84445372 6.73441841 59.31170410
[37] NA -20.00314979 123.70429058 51.06864654 6.43047598 -26.78620936 49.99218244 61.50224193 72.90393286
[46] 47.74473082 93.40113946 24.23157705 91.64188343 100.75378049 90.22672282 49.10556885 45.01707735 -28.75956467
[55] 42.58782906 138.43570309 36.80702958 98.18484589 56.95937526 3.25701298 36.77376741 -57.13274596 130.86377334
[64] 37.74417773 27.39040395 54.30269908 -3.60063702 17.48459480 -3.24157992 -31.55211241 56.90805998 86.31393204
[73] 95.18041151 87.17775672 47.53362842 91.20995329 87.55264706 9.39285802 -20.18253361 -41.13958433 71.54398316
[82] -51.98407283 -4.46249215 68.23219948 22.73161486 NA 54.94947618 -16.02760318 11.13147017 -23.18129214
[91] 80.82314563 8.32210254 -30.63969506 55.53907256 70.59722793 91.42234676 58.26618543 -1.33881045 143.33408823
[100] 8.96361947 -44.52344947 87.89547998 91.22473481 75.08232912 11.55813450 48.47922111 -0.93601496 33.87923014
[109] 50.32103264 7.36905199 74.25089111 36.42348197 117.73715606 58.23406198 -23.16297581 42.99126758 53.74436036
[118] 79.29299388 38.69735295 45.87829756 43.40513285 75.37745460 -33.25107624 43.51654736 12.32215671 44.19736954
[127] 47.40717417 -8.29300039 23.10325522 39.11139531 67.37860719 61.33938505 -36.09153952 81.45155462 63.41171511
[136] 59.60194165 29.26454000 -8.75826158 75.17620628 6.98858771 48.23679725 58.99326555 38.75415672 60.49030246
[145] 96.53599286 -19.89771835 0.05962111 35.27279951 35.92771139 -41.0483337 71.96104669 41.80951143 53.42601644
[154] 95.94866275 24.08387917 NA 79.54290477 -52.08020499 114.00798663 21.16295299 81.23748969 56.23799535
[163] 139.21115441 78.94178791 107.36091299 70.21640201 -14.36520944 15.69281664 -1.05080067 -7.88550666 -11.89179761
[172] 53.93607214 45.39088086 67.72267141 -0.28981337 39.23785049 10.81442625 114.96012693 122.53428693 -45.66809692
[181] 111.49764952 16.90706726 77.85452352 -51.11683332 78.50546064 64.02129892 42.32885469 36.03567390 -0.71419945
[190] 9.23544170 82.73104519 11.36596045 100.57867494 72.09721984 27.29965920 -22.36188198 -28.84295404 24.51721758
[199] -18.56686833 41.21267764 61.15055678 -2.23164474 33.83518407 36.19281926 25.86741445 87.76711246 83.46273615
[208] 17.47412966 78.70810458 50.24784685 15.47952048 65.64621689 9.22711350 -19.27416122 -6.97471296 57.69493051
[217] 5.04580229 -27.21269849 -6.33553272 15.26591282 20.09495943 46.29398444 6.04125774 37.95772065 22.26232483
[226] -32.80745672 39.21479377 23.83563240 -25.75477607 2.66146804 42.65800875 38.69907656 42.53281778 46.66538152
[235] 32.52702378 44.01995102 38.92211457 35.53867807 39.26670620 38.38267489 34.69881693 32.57041727 47.23481662
[244] 47.59676290 47.24578611
> sum(is.na(v))
[1] 4
    
```

Figure 17. Sensitivity of Gaussian 1<sup>st</sup> equation with variation in standard deviation.

```

> x9<-rnorm(230,mean=40,sd=45)
> x10<-rnorm(15,mean=40,sd=5)
> v=c(x9,x10)
> v[v< mean(v)-3*sd(v) | v> mean(v)+3*sd(v)] <- NA #Gaussian Eqn 2
> v
[1] 55.4220163 -2.5936278 -16.6642822 36.9341016 6.0880080 45.5243103 40.1754442 3.8080054 45.6761032
[11] 60.3774192 18.1133825 47.0654328 20.9570015 34.4719768 -68.3862623 48.0549001 4.9505253 -28.8152905 6.0439640
[21] 113.4254321 69.3080579 20.7404335 126.4444656 41.3570430 30.2921152 68.6201796 53.5767099 104.6396365 5.1075656
[31] -36.4190268 75.0362933 -17.1332383 -4.0691281 45.1553530 45.3412994 -16.1215739 52.3968682 -41.8457400 8.2554550
[41] 23.9099394 43.2294645 34.6667081 48.1766365 -9.3583745 94.6652666 65.3005120 106.2787530 4.4277564 69.3923989
[51] 9.4098861 -46.4403371 14.0513569 16.3612140 65.4127078 71.0887433 23.2655054 18.6258683 1.2244419 152.5733107
[61] 34.7782473 19.6429306 -14.1517945 129.7912821 4.8387650 -12.5454678 101.4365525 39.0802086 4.7527112 94.2872117
[71] 0.5614651 55.5998823 5.2792755 -4.3959987 25.4883845 17.3302506 30.3490474 15.4885374 44.2512000 67.8263327
[81] 70.4756650 112.4961192 -40.9882420 7.4640198 61.7318010 72.8173470 -49.7173587 44.7582160 14.8935226 26.8311637
[91] 57.8709321 24.0700478 71.6257836 56.9363148 9.3998848 60.6187896 48.9190361 68.2218742 35.4131633 83.6573456
[101] 73.4940636 -4.6769051 -32.0031148 60.0750041 17.2867271 88.3162684 87.2601889 NA 7.8410947 15.7103815
[111] 14.6925696 -13.3242754 71.3740591 33.1601969 -53.8170438 -16.5543637 -7.9906609 6.7919388 50.5872561 122.8462288
[121] 11.8401486 9.9210465 69.3023100 -88.0466115 18.5009451 23.1012545 -2.2551452 42.3700687 49.2607813 87.6588359
[131] 46.3144404 1.7485667 -38.6726127 54.4227870 66.1567389 117.8207481 56.0408366 -25.9985560 68.8770226 93.4995950
[141] 137.0232307 30.5403781 32.5060718 -2.3638694 128.7620506 64.8282124 43.9095920 -27.8526446 128.0233133 54.8228024
[151] 47.0291371 -16.1404512 64.4899266 4.5682421 38.2701934 19.4683476 31.7924935 6.4494694 49.6379623 15.9316833
[161] 21.4542191 -41.7235297 25.2198873 -73.8894880 -76.1539413 -64.3802832 36.2905116 1.1022525 46.7695107 108.0784005
[171] 92.9020655 -45.5421194 95.5186453 51.1608519 82.5598221 28.4041956 3.4287698 62.3828105 88.0786201 83.1795500
[181] 63.7956762 40.5927677 -0.5919659 -45.0593531 26.9563472 57.0686733 37.3803973 71.7441066 104.2674301 21.9282165
[191] 44.9920590 54.3167210 54.8720545 -52.0100366 63.7197220 38.2078991 61.2456523 9.8755302 119.2169590 69.6266953
[201] 28.1726907 -2.5440946 -7.4648953 69.1477550 8.6702657 -18.5746779 50.1558257 42.2589552 59.6716937 129.3575447
[211] 52.1857321 69.8380636 36.2786637 120.5603905 24.1811386 60.1419984 -52.2249802 -25.6127521 109.6401894 15.2353117
[221] -39.2704180 69.6527402 72.7163004 83.9769398 62.8285959 19.3448041 42.6829576 100.2652100 55.0088426 45.9475374
[231] 37.1959997 43.9934572 38.1647470 31.9432194 50.8032332 39.7858865 37.8601283 29.5426788 34.9634230 36.2750891
[241] 41.8387092 40.3503735 45.4844198 39.4928750 41.6243325
> sum(is.na(v))
[1] 1
    
```

Figure 18. Sensitivity of Gaussian 2<sup>nd</sup> equation with variation in standard deviation.

```

> x9<-rnorm(230,mean=40,sd=45)
> x10<-rnorm(15,mean=40,sd=5)
> v=c(x9,x10)
> v[abs(quantile(v,.50)-3*exp(Gam(Top(Cabs(v)-quantile(v,.50))+1))/length(v)) |
+ abs(quantile(v,.50)+3*exp(Gam(Top(Cabs(v)-quantile(v,.50))+1))/length(v)) <- NA #New Eqn
> v
[1] 83.0003220 -16.0160947 82.5625977 76.1778598 79.8875513 91.6410314 36.0199790 15.9509939 -8.8210540 58.0028093
[11] 82.8523173 65.5849490 11.9178802 NA NA 23.9512329 14.4381314 NA 71.4914007 62.5490353
[21] NA 99.6061421 31.8317139 35.1669650 21.6033049 63.7331570 69.1193972 10.7551820 83.5671381 17.1318838
[31] 48.1602579 20.4586349 36.6576533 71.8587294 22.9007857 38.0963344 NA -1.0439209 18.9099837 53.3218875
[41] -1.4944356 65.5665808 -7.9798436 39.6161539 28.7400734 -16.4411058 39.8888055 -4.6400650 NA NA
[51] 34.7444625 27.5807926 60.2755944 -0.8917742 14.3371179 42.8309166 52.4700779 -13.5897983 35.8393106 10.8892945
[61] 76.4496730 83.8357013 17.8561561 48.0159719 24.8257803 41.9617676 42.3077437 15.1704014 -18.4229346 77.4861925
[71] NA 71.8795124 44.137246 NA NA 47.4805128 NA 05.7940998 52.9457840 106.8130387 69.3514946
[81] 48.4299133 21.8392910 9.3945479 107.8667726 22.1531140 103.7189918 NA NA -20.3966670 84.1711693
[91] 58.6474210 -4.6014210 37.9682874 -16.4594135 NA 28.5865715 68.0875804 82.0743905 NA 87.0652670
[101] 98.6288034 16.9318272 3.6458591 101.3947763 NA -26.8394793 72.0248214 49.2719955 24.7925790 62.5984231
[111] NA 87.2840858 92.5656750 24.2577259 55.7991547 82.2362858 105.3054354 43.4579984 63.4654339 62.6912436
[121] 98.5402645 88.3351028 23.9437836 70.5973969 49.8471972 80.5539111 4.3917542 68.7709954 14.77130526 96.1018252
[131] 7.5490041 19.0785183 NA 3.4169198 22.8305457 11.9752417 -13.9284695 NA 101.4742665 -3.3969321
[141] 76.5847567 63.1593826 8.2225395 -1.8703404 -19.3107980 64.2051937 46.0840086 81.6083344 69.0981956 27.7943635
[151] 51.7494860 10.8718594 86.6117472 68.2177489 -17.1829321 NA -23.2998131 35.1908955 61.7541475 35.1318602
[161] 17.0571647 NA NA -3.0278321 59.4350300 54.1674750 68.5944707 19.1900680 33.5992706 5.4725860 80.7529181
[171] 101.1292426 94.3087859 27.0525996 -22.3714397 NA 9.1677918 55.0051134 17.4631891 -1.1631884 12.7269550
[181] 61.6501408 96.3234251 78.2735465 9.2309587 78.1874383 72.0950887 84.9382541 94.4223395 30.4397519 41.0160731
[191] 104.6514356 20.9160341 51.7797152 45.0366587 41.1551604 62.1301696 48.1773267 -9.9143703 56.8180087 49.8754899
[201] 18.0220449 58.8329291 19.2568349 NA 57.3627202 NA -23.7864201 -0.5682756 61.0880842 70.3898822
[211] 50.7851336 -7.2874416 NA NA 43.1987587 NA 54.6782281 3.1641006 11.1046706 23.1057661
[221] 79.4403966 57.3017750 23.6805500 31.8881267 61.8928224 20.1559579 33.7068663 2.6141449 NA 4.0862699
[231] 42.6687635 43.3605715 38.0756078 37.9538322 31.7028694 46.3859032 42.8689759 39.7207347 36.6704813 39.4206104
[241] 27.6898790 39.5040903 37.3056026 37.0030377 35.9573588
> sum(is.na(v))
[1] 27
    
```

Figure 19. Sensitivity of the New Equation with variation in standard deviation.

Appendix 5. Outlier Detection Simulation for Very Large Data Sets with Standard Deviation Variation

```

> set.seed(45)
> x5<-rnorm(500,mean=20,sd=10)
> x6<-rnorm(55,mean=20,sd=5)
> v=c(x5,x6)
> v[<v<quantile(v,.25)-1.5*IQR(v) | v>quantile(v,.75)+1.5*IQR(v)] <- NA #Gaussian Eqn 1
> v
[1] 23.40799691 12.96659699 16.20462269 12.53952563 11.01892670 16.65205892 14.98621847 18.25464303 38.09037403 17.69895026
[11] 8.69581782 22.15988887 32.32237285 36.09358707 24.0150627 17.27015966 19.63847658 18.49688766 NA 3.475042023
[21] 8.64854898 22.27670169 18.16681463 15.86481375 15.62404717 19.73815646 11.40165823 21.66544577 34.75490730 21.95422914
[31] 21.59421791 12.79806715 10.64497457 22.85432305 12.60764853 24.29148980 NA 6.66596671 38.60095346 22.45969923
[41] 12.54010577 5.15862014 22.22048499 24.77827598 27.31198508 21.72105129 31.86691988 16.49590601 31.4757410 33.50084011
[51] 31.6149396 22.02345974 21.11490608 4.72009670 20.5701212 41.74060069 19.29045102 32.92996945 7.6875006 18.83489111
[61] 19.28305942 -1.58505466 40.19194963 19.49870616 17.19786754 23.17837757 10.31096955 14.99657662 10.39076002 4.09953057
[71] 23.75734666 30.29198490 32.26231367 30.48394594 21.67413965 31.37998962 30.56725490 13.19841289 6.62610364 1.96898126
[81] 27.00977404 -0.44090507 10.11944619 26.27382211 16.16258108 NA 23.32210582 7.54942152 22.47366004 5.95971286
[91] 29.07181014 12.96046723 4.30228999 23.45312724 26.79938398 31.42718817 24.05915232 10.81359768 42.96313072 13.10302655
[101] 1.21701123 30.6434000 31.38327440 27.79607314 13.67958544 21.88427136 10.90310779 18.63982892 22.29356281 12.74867822
[111] 27.61130914 19.20521822 27.27492357 24.05203377 9.96378315 20.66472613 23.05430230 28.73177642 19.71052288 21.95268835
[121] 20.75669619 27.86165658 3.72198306 20.78145497 13.84936816 20.93274879 21.64603870 9.26822214 16.24516783 19.80253229
[131] 26.08413493 24.7208557 3.09076900 29.21145658 25.20260336 24.35598703 17.61434222 9.16483076 27.81693473 12.66413060
[141] 21.83039939 24.22072568 19.72314594 24.55340055 32.56355397 6.68939592 11.12436025 18.94951100 19.09504697 1.98925925
[151] 27.10245482 20.40211365 22.9835921 32.43303617 16.46308426 NA 28.78731217 -0.46226778 36.44621925 15.81398955
[161] 29.16388600 23.6084341 42.04692280 28.65373065 34.96909178 26.71475600 7.91884235 14.59840370 10.87759985 9.35876652
[171] 8.46848942 23.09604992 21.19797352 26.16059365 11.04670814 2.39158878 13.51431695 36.65780598 38.34095285 0.96264513
[181] 35.88836656 14.86823717 38.41211634 -0.24818518 28.55676903 25.33806643 20.51752326 19.11903864 11.07240012 13.16341349
[191] 29.49578782 13.63688010 33.46292776 27.13271552 17.1770204 6.14180401 4.70156569 16.55938169 6.98514037 20.26948392
[201] 24.70012373 10.61519006 18.63004090 19.15395984 16.85942543 30.61489139 29.65838581 14.99365104 28.60180102 22.2779930
[211] 14.55100455 25.6915931 13.16158078 6.82796417 9.56117350 23.93220678 12.23240051 5.06384478 9.70321495 14.50353618
[221] 15.57665765 21.39865621 21.45361283 15.56160114 05.08294411 8.82055517 6.828551006 16.40791831 5.38782754 11.70254845
[231] 25.31601750 17.39815312 25.06563556 33.33076303 5.05404756 28.03900205 17.84422914 11.0735613 18.5341240 16.76534977
[241] 9.39763386 5.14083454 34.46963324 35.19352580 34.49157223 23.42711473 10.53474937 4.70793730 19.31868924 12.46400177
[251] 24.31962627 22.2762450 20.03898760 11.95733453 21.26135627 24.52831537 15.13630721 21.57009617 15.76822255 18.77155040
[261] -4.08583606 21.78997780 12.21122784 4.7071322 12.45421422 36.31676268 26.51290176 15.72009634 39.20988124 20.3016512
[271] 17.84269226 26.36003991 23.01704663 34.36436367 12.24612568 3.01799404 27.78584296 7.30372482 10.20686041 21.14630067
[281] 24.18695342 7.52839312 27.74893961 1.81205778 12.94565667 16.42442099 20.7165879 18.81482403 21.81700033 9.0347223
[291] 32.14783702 25.62233600 34.72861177 12.0905698 26.53164420 13.20219291 0.79103620 14.23363486 14.74693645 25.64726841
[301] 26.90860961 16.28122342 15.25019295 11.38320932 NA 18.83961650 15.47620679 7.96626789 39.95361824 12.18639222
[311] 8.32322938 33.65256722 19.79560191 12.16726916 32.06382482 11.23588114 23.46664051 12.28428344 10.13422252 16.77519656
[321] 14.96227917 17.85534387 14.55300830 20.94471111 26.18362949 26.77237000 36.11024871 2.00261289 12.76978218 24.89298911
[331] 27.29274378 0.06280919 21.05738133 14.42078281 17.07359193 23.97131825 16.46001062 27.02795191 23.76362550 13.19997440
[341] 24.58195282 98.200803 26.27152760 18.98070296 19.70162326 17.44312525 10.0719888 3.99930783 24.4611202 14.95260602
[351] 30.73694853 30.50226420 NA 12.85357660 21.26897367 14.37612657 8.15016101 26.97201312 18.48004375 -0.84823195
[361] 7.43236362 9.33540870 12.62043085 22.35272357 38.41027308 13.74225523 13.31578812 26.51162444 NA 15.2243225
[371] 16.24472322 10.60996774 20.52668193 22.05795140 30.59085241 21.40320121 11.49968148 2.51719717 23.20506377 25.81260865
[381] 37.29349958 23.56463036 5.33365422 26.41711612 31.88879090 41.56071793 17.89786181 18.33468263 10.58580680 19.72490014
[391] 25.51738053 20.86879822 4.92163453 39.56073629 23.29395610 21.56203046 7.52434417 25.44220591 12.12627602 19.61559852
[401] 15.43741057 18.17610967 12.54432654 22.14176939 14.65148518 15.87871535 1.83921563 16.71553052 -5.30877511 -8.11969695
[411] -3.19561849 19.17566925 11.35605612 21.50433570 35.12853344 31.75601456 0.99064014 32.33747674 22.48018932 25.47732824
[421] 17.42315457 11.87305996 24.97395789 30.68413780 29.71545756 25.28837249 20.13172616 10.97956313 1.09792153 17.10141050
[431] 23.79303851 19.41786606 27.05424592 34.28165113 15.98404812 21.10112242 23.18149356 33.30490100 -0.44667480 25.27104933
[441] 19.60175536 24.72125607 13.30567337 37.60376867 15.47259895 17.49170905 10.54575675 9.45224550 26.47727888 13.03783683
[451] 6.98340492 22.25685015 20.50190040 24.37148748 39.85722994 22.7094047 26.63068079 19.17303639 37.90230899 16.48649747
[461] 24.47599965 -0.49444004 5.41938843 35.47559765 14.49676680 2.38435155 26.63068079 12.203899 27.27028897 27.7265328
[471] 15.40995647 20.59621279 33.39226889 23.33529835 21.32167498 14.39199930 27.98691437 16.32949402 3.88643871 41.60646637
[481] 19.57177308 15.72025659 -0.91464239 9.92684605 12.55017818 23.67741840 20.70074700 30.96883968 18.98574999 23.2486507
[491] 29.5562710 7.55197896 29.45835505 28.03952441 28.86390028 31.47587475 19.11555088 14.65577642 9.15087689 24.00062429
[501] 24.76136859 22.8427211 16.87976446 29.7296337 28.6725859 18.21680366 17.15979238 7.44298979 23.49904452 22.50543837
[511] 29.30138972 16.62209470 19.09241208 19.46299611 17.9559227 22.6370145 23.25488859 14.36093513 24.8409312 17.53553486
[521] 20.90669532 17.82873717 19.62862814 23.53985882 18.10008730 19.78848159 11.63517121 15.43956434 17.65666486 21.48043194
[531] 15.38953826 22.84073120 14.66890626 19.95735043 18.74956371 13.72876602 19.98764505 16.07111834 28.51953234 10.54072760
[541] 19.41605139 18.62008807 22.25284049 15.45646953 17.14856856 20.31454629 21.38556421 14.04780019 19.53770118 16.76547716
[551] 24.04996366 24.87063347 17.53957292 20.97952543 18.32508670
> sum(is.na(v))
[1] 8

```

Figure 20. Sensitivity of Gaussian 1<sup>st</sup> equation with vari- in standard deviation.

```

> set.seed(45)
> x5<-rnorm(500,mean=20,sd=10)
> x6<-rnorm(55,mean=20,sd=5)
> v=c(x5,x6)
> v[<v<mean(v)-3*sd(v) | v>=mean(v)+3*sd(v)] <- NA #Gaussian Eqn 2
> v
[1] 23.40799691 12.96659699 16.20462269 12.53952563 11.01892670 16.65205892 14.98621847 18.25464303 38.09037403 17.69895026
[11] 8.69581782 22.15988887 32.32237285 36.09358707 24.0150627 17.27015966 19.63847658 18.49688766 NA 3.475042023
[21] 8.64854898 22.27670169 18.16681463 15.86481375 15.62404717 19.73815646 11.40165823 21.66544577 34.75490730 21.95422914
[31] 21.59421791 12.79806715 10.64497457 22.85432305 12.60764853 24.29148980 47.33983845 6.66596671 38.60095346 22.45969923
[41] 12.54010577 5.15862014 22.22048499 24.77827598 27.31198508 21.72105129 31.86691988 16.49590601 31.4757410 33.50084011
[51] 31.6149396 22.02345974 21.11490608 4.72009670 20.5701212 41.74060069 19.29045102 32.92996945 7.6875006 18.83489111
[61] 19.28305942 -1.58505466 40.19194963 19.49870616 17.19786754 23.17837757 10.31096955 14.99657662 10.39076002 4.09953057
[71] 23.75734666 30.29198490 32.26231367 30.48394594 21.67413965 31.37998962 30.56725490 13.19841289 6.62610364 1.96898126
[81] 27.00977404 -0.44090507 10.11944619 26.27382211 16.16258108 NA 23.32210582 7.54942152 22.47366004 5.95971286
[91] 29.07181014 12.96046723 4.30228999 23.45312724 26.79938398 31.42718817 24.05915232 10.81359768 42.96313072 13.10302655
[101] 1.21701123 30.6434000 31.38327440 27.79607314 13.67958544 21.88427136 10.90310779 18.63982892 22.29356281 12.74867822
[111] 27.61130914 19.20521822 27.27492357 24.05203377 9.96378315 20.66472613 23.05430230 28.73177642 19.71052288 21.95268835
[121] 20.75669619 27.86165658 3.72198306 20.78145497 13.84936816 20.93274879 21.64603870 9.26822214 16.24516783 19.80253229
[131] 26.08413493 24.7208557 3.09076900 29.21145658 25.20260336 24.35598703 17.61434222 9.16483076 27.81693473 12.66413060
[141] 21.83039939 24.22072568 19.72314594 24.55340055 32.56355397 6.68939592 11.12436025 18.94951100 19.09504697 1.98925925
[151] 27.10245482 20.40211365 22.9835921 32.43303617 16.46308426 -8.50789299 28.78731217 -0.46226778 36.44621925 15.81398955
[161] 29.16388600 23.6084341 42.04692280 28.65373065 34.96909178 26.71475600 7.91884235 14.59840370 10.87759985 9.35876652
[171] 8.46848942 23.09604992 21.19797352 26.16059365 11.04670814 2.39158878 13.51431695 36.65780598 38.34095285 0.96264513
[181] 35.88836656 14.86823717 38.41211634 -0.24818518 28.55676903 25.33806643 20.51752326 19.11903864 11.07240012 13.16341349
[191] 29.49578782 13.63688010 33.46292776 27.13271552 17.1770204 6.14180401 4.70156569 16.55938169 6.98514037 20.26948392
[201] 24.70012373 10.61519006 18.63004090 19.15395984 16.85942543 30.61489139 29.65838581 14.99365104 28.60180102 22.2779930
[211] 14.55100455 25.6915931 13.16158078 6.82796417 9.56117350 23.93220678 12.23240051 5.06384478 9.70321495 14.50353618
[221] 15.57665765 21.39865621 21.45361283 15.56160114 05.08294411 8.82055517 6.828551006 16.40791831 5.38782754 11.70254845
[231] 25.31601750 17.39815312 25.06563556 33.33076303 5.05404756 28.03900205 17.84422914 11.0735613 18.5341240 16.76534977
[241] 9.39763386 5.14083454 34.46963324 35.19352580 34.49157223 23.42711473 10.53474937 4.70793730 19.31868924 12.46400177
[251] 24.31962627 22.2762450 20.03898760 11.95733453 21.26135627 24.52831537 15.13630721 21.57009617 15.76822255 18.77155040
[261] -4.08583606 21.78997780 12.21122784 4.7071322 12.45421422 36.31676268 26.51290176 15.72009634 39.20988124 20.3016512
[271] 17.84269226 26.36003991 23.01704663 34.36436367 12.24612568 3.01799404 27.78584296 7.30372482 10.20686041 21.14630067
[281] 24.18695342 7.52839312 27.74893961 1.81205778 12.94565667 16.42442099 20.7165879 18.81482403 21.81700033 9.0347223
[291] 32.14783702 25.62233600 34.72861177 12.0905698 26.53164420 13.20219291 0.79103620 14.23363486 14.74693645 25.64726841
[301] 26.90860961 16.28122342 15.25019295 11.38320932 15.01629126 18.83961650 15.47620679 7.96626789 39.95361824 12.18639222
[311] 8.32322938 33.65256722 19.79560191 12.16726916 32.06382482 11.23588114 23.46664051 12.28428344 10.13422252 16.77519656
[321] 14.96227917 17.85534387 14.55300830 20.94471111 26.18362949 26.77237000 36.11024871 2.00261289 12.76978218 24.89298911
[331] 27.29274378 0.06280919 21.05738133 14.42078281 17.07359193 23.97131825 16.46001062 27.02795191 23.76362550 13.19997440
[341] 24.58195282 98.200803 26.27152760 18.98070296 19.70162326 17.44312525 10.0719888 3.99930783 24.4611202 14.95260602
[351] 30.73694853 30.50226420 NA 12.85357660 21.26897367 14.37612657 8.15016101 26.97201312 18.48004375 -0.84823195
[361] 7.43236362 9.33540870 12.62043085 22.35272357 38.41027308 13.74225523 13.31578812 26.51162444 -8.45480255 15.2243225
[371] 16.24472322 10.60996774 20.52668193 22.05795140 30.59085241 21.40320121 11.49968148 2.51719717 23.20506377 25.81260865
[381] 37.29349958 23.56463036 5.33365422 26.41711612 31.88879090 41.56071793 17.89786181 18.33468263 10.58580680 19.72490014
[391] 25.51738053 20.86879822 4.92163453 39.56073629 23.29395610 21.56203046 7.52434417 25.44220591 12.12627602 19.61559852
[401] 15.43741057 18.17610967 12.54432654 22.14176939 14.65148518 15.87871535 1.83921563 16.71553052 -5.30877511 -8.11969695
[411] -3.19561849 19.175
```

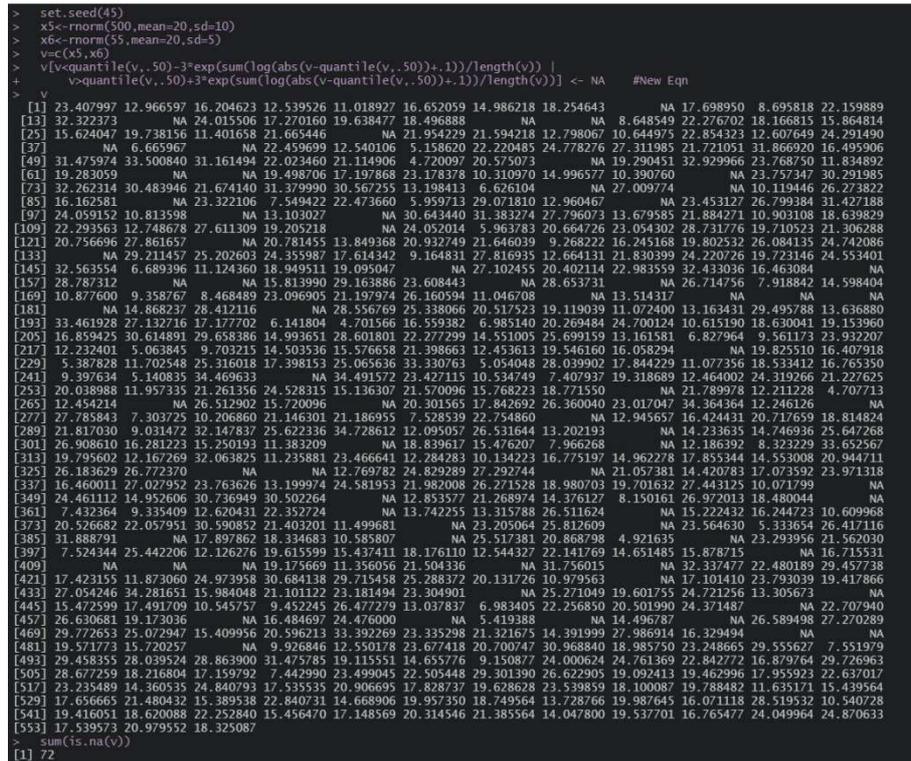


Figure 22. Sensitivity of the New Equation with variation in standard deviation.

References

[1] I. Ben-Gal, "Outlier detection," in Data mining and knowledge discovery handbook, Springer, 2005, pp. 131-146.

[2] P. L. Clark, "Number theory: A contemporary introduction", 2012.

[3] C. E. Shannon, A mathematical theory of communication, vol. 27, Bell System Technical Journal, 1948, pp. 379-423.

[4] D. Papadopolus, T. Palpanas, D. Gonupulos and V. Kalogeraki, "Distributed deviation detection in sensor networks", vol. 32, Acm sigmod record, 2003, pp. 77-82.

[5] J. Orsborne and A. Overbay, "The power of outliers (and why researchers should always check for them)", vol. 9, Practical Assessment, Research and Evaluation, 2004, p. 6.

[6] X. Li and J. Han, "Mining approximate top k subspace anomalies in multidimensional time series data", in VDLBD, 2007, pp. 447-458.

[7] V. Ganti, R. Ramakrishnan, J. Gehrke, A. Powell and J. French, "Clustering large datasets in arbitrary metric spaces", in Proceedings 15th International Conference on Data Engineering (Cat. No. 99CB36337), pp. 502-511.

[8] S. D. Bay and M. Schwabacher, "Mining distance-based outliers in near linear time with randomization and a sample pruning rule", in Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, 2003, pp. 29-38.

[9] E. Eskin, A. Arnold, M. Prerav, L. Portnoy and S. Stolfo, "A geometric framework for unsupervised anomaly

detection", in Applications of data mining in computer security. Springer, 2002, pp. 77-101.

[10] J. Han and M. Kamber, "Data mining concepts and techniques", San Francisco: morgan kaufmann publishers.

[11] V. Barnett, "The ordering of multivariate data", vol. 139, Journal of royal statistical society Series A (General), 1976, pp. 318-344.

[12] P. J. Rousseeuw and M. Hubert, "Robust statistics for outlier detection", vol. 1, Wiley interdisciplinary reviews. Data mining and knowledge discovery, 2011, pp. 73-79.

[13] C. C. Aggarwals, J. Han, J. Wang and P. S. Yu, A framework for projected clustering of high dimensional data streams", vol. 30, in Proceedings of the Thirtieth international conference on very large databases, 2004, pp. 852-863.

[14] P. C. Wu, "The Central Limit Theorem and comparing means, trimmed means, one-step M-estimators and modified one-step M-estimators under non-normality", University of Southern California, 2002.

[15] A. Biswas and A. Bisaria, "A test of normality from allegorizing the bell curve or the gaussian probability distribution as memoryless and depthless like a black hole", vol. 14, Applied Mathematics Sciences, 2020, pp. 349-359.

[16] R. Lugannani and S. Rice, "Saddle point approximation for the distribution of the sum of independent random variables", vol. 12, Advances in applied probability, 1980, pp. 475-490.

[17] C. Leys, C. Ley, O. Klein, P. Bernard and L. Licata, Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median", vol. 49, Journal of experimental social psychology, 2013, pp. 764-766.

- [18] P. J. Rouseeuw and B. C. Van Zomeren, "Unmasking multivariate outliers and leverage points", vol. 85, *Journal of the American Statistics association*, 1990, pp. 633-639.
- [19] V. L. Sourd, "Performance measurement for traditional investment", vol. 58, *Financial Analysis Journal*, 2007, pp. 36-52.
- [20] P. V. Hippel, "Mean, median and skew: correcting a textbook rule", vol. 13, *Journal of statistics Education*, 2005.
- [21] E. M. Knorr and R. T. Ng, "Finding intensional knowledge of distance-based outliers", vol. 99, in *Vldb*, 1999, pp. 211-222.
- [22] W. Dixon, "Processing data for outliers", vol. 9, *Biometry*, 1953, pp. 74-89.
- [23] F. Angiulli and C. Pizzuti, "fast outlier detection in high dimensional spaces, principals of data mining and knowledge discovery", 2002.
- [24] B. Troon, "Estimating average variation about the population mean using geometric measure of variation", 2020.
- [25] T. Li, H. Fan, J. Garcia and J. M. Corchado, "Second order statistics analysis and comarison between arithmetic and geometric average fusion: Application to multi-sensor target tracking", vol. 51, *Information Fusion*, 2019, pp. 233-243.
- [26] V. Barnett and T. Lewis, "Outliers in statistical data, Wiley series in Probability and Mathematical statistics. Applied Probability and Statistics 1984.