**Open Journal of Statistics, 2014, 4, 536-561**
Published Online August 2014 in SciRes. http://www.scirp.org/journal/ojs
http://dx.doi.org/10.4236/ojs.2014.47051

# Comparison of Outlier Techniques Based on Simulated Data

**Adaku C. Obikee[1*], Godday U. Ebuh[2], Happiness O. Obiora-Ilouno[1]**

[1]Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Nigeria
[2]Monetary & Policy Department, Central Bank of Nigeria, Abuja, Nigeria
Email: *pobikeeadaku@yahoo.com, ablegod007@yahoo.com, obiorailounoho@yahoo.com

Received 5 June 2014; revised 8 July 2014; accepted 18 July 2014



## Abstract

**This research work employed a simulation study to evaluate six outlier techniques: *t*-Statistic, Modified *Z*-Statistic, Cancer Outlier Profile Analysis (COPA), Outlier Sum-Statistic (OS), Outlier Robust *t*-Statistic (ORT), and the Truncated Outlier Robust *t*-Statistic (TORT) with the aim of determining the technique that has a higher power of detecting and handling outliers in terms of their *P*-values, true positives, false positives, False Discovery Rate (FDR) and their corresponding Receiver Operating Characteristic (ROC) curves. From the result of the analysis, it was revealed that OS was the best technique followed by COPA, *t*, ORT, TORT and *Z* respectively in terms of their *P*-values. The result of the False Discovery Rate (FDR) shows that OS is the best technique followed by COPA, t, ORT, TORT and *Z*. In terms of their ROC curves, *t*-Statistic and OS have the largest Area under the ROC Curve (AUC) which indicates better sensitivity and specificity and is more significant followed by COPA and ORT with the equal significant AUC while *Z* and TORT have the least AUC which is not significant.**

## Keywords

**Area under the ROC Curve, Reference Line, Sensitivity, Specificity, *P*-Value, False Discovery Rate (FDR), Simulation**

## 1. Introduction

In statistics, an outlier is an observation that is numerically distant from the rest of the data. [1] Grubbs (1969) defined an outlier as an observation that appears to deviate markedly from other members of the sample in which it occurs. [2] Hawkins (1980) formally defined the concept of an outlier as "an observation which de-

*Corresponding author.

**How to cite this paper:** Obikee, A.C., Ebuh, G.U. and Obiora-Ilouno, H.O. (2014) Comparison of Outlier Techniques Based on Simulated Data. *Open Journal of Statistics*, **4**, 536-561. http://dx.doi.org/10.4236/ojs.2014.47051

viates so much from the other observations so as to arouse suspicions that it was generated by a different mechanism". Outliers are also referred to as abnormalities, discordants, deviants, or anomalies in data mining and statistics literature [3] (Aggarwal, 2005). Outliers can also be defined in a closed bound: For example, if $Q_1$ and $Q_3$ are the lower and upper quartiles of a sample, then one can define an outlier to be any observation outside the range: $\left[Q_1 - k\left(Q_3 - Q_1\right), Q_3 + k\left(Q_3 - Q_1\right)\right]$ for some constant $k$ [4] (Barnett and Lewis, 1994). Outliers can occur by chance in any distribution, but they are often indicative either of measurement error or that the population has a heavy-tailed distribution. Outliers provide interesting case studies. They should always be identified and discussed. They should never be ignored, or "swept under the rug". In any scientific research, full disclosure is the ethical approach, including a disclosure and discussion of the outliers.

In many analyses, outliers are the most interesting things. Outliers often provide valuable insight into particular observations. Knowing why an observation is an outlier is very important. For example, outlier identification is a key part of quality control. The box plot and the histogram can also be useful graphical tools in checking the normality assumption and in identifying potential outliers. While statistical methods are used to identify outliers, non-statistical theory (subject matter) is needed to explain why outliers are the way that they are.

In sampling of data, some data points will be farther away from the sample mean than what is deemed reasonable. This can be due to incidental errors or flaws in the theory that generated an assumed family of probability distributions or it may be that some observations are far from the center of the data. Outlier points can therefore indicate faulty data, erroneous procedures, or areas where a certain theory might not be valid. Outliers can occur by chance in any distribution, but they are often indicative *either* of measurement error or that the population has a heavy-tailed distribution. In the former case one wishes to discard them or use statistics that are robust to outliers, while in the latter case they indicate that the distribution has high kurtosis.

Hence, this study is set out to evaluate six different outlier techniques using their *P*-values, true positives, false positives, FDRs and their corresponding Receiver Operating Characteristics ROC Curves using a simulated data.

Researchers that have similar work in this regards include [5] Dudoit *et al.* (2002), [6] Troyanskaya *et al.* (2002), [7] Tomlins *et al.* (2005), [8] Efron *et al.* (2001), [9] Iglewicz and Hoaglin (2010), [10] Lyons *et al.* (2004), [11] Tibshirani and Hastie (2006), [12] Benjamini and Hochberg(1995), [13] Wu (2007), [14] June (2012), [15] Fonseca (2004), [16] MacDonald and Ghosh (2006), [17] Jianhua (2008), [18] Heng(2008), [19] Ghosh (2009), [20] Lin-An *et al.* (2010), [21] Ghosh (2010), [22] Filmoser *et al.* (2008) and [23] Keita *et al.* (2013)

## 2. Method of Analysis

The six outlier methods include: The Modified *Z*-statistic, *t*-Statistic, OS, COPA, ORT and TORT.

This paper considers a 2-class data for detecting outliers. Let $x_{ij}$ be the expression values for the normal group for $i = 1, 2, \cdots, n_1$ and $j = 1, 2, \cdots, p$ the number of sample groups and let $y_{ij}$ be the expression values for the disease group and $i = 1, \cdots, n_2$ and $j = 1, 2, \cdots, p$. Where $n_1 + n_2 = n$.

The standard *Z*-statistic for 1 sample test is

$$Z = \frac{y_i - \bar{y}}{s} \tag{1}$$

Iglewicz and Hoaglin (2010) recommend using the modified *Z*-score

$$MZ_i = \frac{0.6745\left(y_i - \tilde{y}\right)}{\text{MAD}} \quad i = 1, 2, 3, \cdots, n \tag{2}$$

With MAD denoting the median absolute deviation, $y_i$ are the observed values and $\tilde{y}$ denoting the median $Med_i$.

These authors recommended that modified *Z*-scores with an absolute value of greater than 3.5 be labeled as potential outliers. *i.e.*

$$MZ_i = \left| \frac{y_i - Med_i}{Mad_i} \cdot 0.6745 \right| > 3.5 \tag{3}$$

The *t*-Statistic for a two sample test by Dudoit *et al.* (2002) and Troyanskaya *et al.* (2002) is given as:

$$t = \frac{\bar{Y}_i - \bar{X}_i}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad i = 1, 2, 3, \cdots, n \tag{4}$$

Here, $\bar{x}_i$ and $\bar{y}_i$ are the sample means for *i* in the normal group and the disease group respectively. The denominator is the pooled standard deviation for the variable *i*.

Tomlins (2005) defined the COPA statistic, which is the $r^{th}$ percentile of standardized samples in the disease group. The COPA statistic has the formula:

$$C_i = \frac{q_r\left(y_{ij} - Med_i\right)}{Mad_i} \qquad i = 1, 2, 3, \cdots, n \text{ and } j = 1, \cdots, p \tag{5}$$

$$\text{for } \left(y_{ij} : 1 \geq i \geq n_2\right)$$

where $q_r$ is the $r^{th}$ percentile of the data, $Medi_i$ is the median of all values for *i*, and $Madi_i$ is the median absolute deviation of all expressions for *i*, *n* is the number observations and *p* is the number of sample group. The choice of *r* is subjective. Obviously, the COPA statistic $C_i$ only utilizes a single value $\left(y_i : 1 \leq i \leq n_2\right)$.

According to Tibshirirani and Hastie (2006),

$$\text{OS}_i = \frac{\sum_{i \varepsilon O_i}\left(y_{ij} - Med_i\right)}{Mad_i} \tag{6}$$

where $O_i \varepsilon \left(Q_1 - IQR, Q_3 + IQR\right)$, $i = 1, 2, 3, \cdots, n$ and $j = 1, \cdots, p$ $Q_1$, $Q_3$ and $IQR$ are the first quartile, third quartile and the interquartile range of all expressions for *i* respectively, *n* is the number observations and *p* is the number of sample groups. Outlier-sum statistic defines outliers in the disease group based on the pooled sample for *i*

Accordingly, Wu (2007) defined Outlier Robust *t*-Statistic ORT as:

$$\text{ORT}_i = \frac{\sum_{i \varepsilon_{Oi}}\left(Y_{ij} - Med_i^c\right)}{Median\left\{\left|X_{ij} - Med_i^c\right|_{j \leq n_1} \left|Y_{ij} - Med_i^d\right|_{j \leq n_2}\right\}} \qquad i = 1, 2, 3, \cdots, n \text{ and } j = 1, \cdots, p \tag{7}$$

where $Med_i^c = Median\left(X_{ij} : 1 \leq j \leq n_1\right)$, $Med_i^d = Median\left(Y_{ij} : 1 \leq j \leq n_2\right)$ and $O_i \varepsilon \left(Q_1 - \text{IQR}, Q_3 + \text{IQR}\right)$.

The statistic ORT concentrates on the outlier set $O_i$. However, it uses all the values from disease group.

According to June (2012), TORT is given as:

$$\text{TORT}_i = \frac{\sum_{i \varepsilon_{Oi}}\left(Y_{ij} - Med_i^c\right)}{Median\left\{\left|X_{ij} - Med_i^c\right|_{i \leq n_1} \left|Y_{ij} - Med_i^{oi}\right|_{i \leq n_2}\right\}} \qquad i = 1, 2, 3, \cdots, n \text{ and } j = 1, \cdots, p \tag{8}$$

where $Med_i^c = Median\left(X_{ij} : 1 \leq j \leq n_1\right)$, $Med_i^o = Mediano_i$ and $O_i \varepsilon \left(Q_1 - \text{IQR}, Q_3 + \text{IQR}\right)$

The false discovery rate can be calculated using the modified formula:

$$\text{FDR} = \frac{\text{FP}}{\text{FP+TP}} \tag{9}$$

where FP and TP are the False Positive (Specificity) and the True Positive (Sensitivity).

**Quartiles:** The 1st quartiles $Q_1$, 2nd quartiles, the 3rd quartiles $Q_3$ and the Interquartile Range IQR of each of the sample the simulated data were calculated for analysis. The quartiles can be calculated using the modified formula:

$$L_y = n \cdot \frac{y}{100} \tag{10}$$

where $L_y$ is the required quartile, *y* is the percentile of the require quartile and *n* is the number of observation.

- First quartile (designated $Q_1$) = lower quartile = splits lowest 25% of data = 25th percentile.

- Second quartile (designated $Q_2$) = median = cuts data set in half = 50th percentile.
- Third quartile (designated $Q_3$) = upper quartile = splits highest 25% of data, or lowest 75% = 75th percentile. The difference between the upper and lower quartiles is called the interquartile range IQR.

**Area under the ROC Curve (AUC):** The area under the ROC curve AUC can be estimated using the modified formula:

$$\text{AUC} = \phi\left(\frac{\mu_x - \mu_y}{\sqrt{\sigma_x^2 + \sigma_y^2}}\right) \tag{11}$$

where $\mu_x$ and $\mu_y$ are the mean of the specificities and the sensitivities. $\sigma_x^2$ and $\sigma_y^2$ are the standard deviations of the specificities and sensitivities.

a) The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

b) **Null Hypothesis**: Significant/True Area under the ROC Curve (AUC) = 0.5.

## 3. Data Simulation and Application

**Table 1** is random numbers generated from a normal distribution with parameters-sample size $n$ = 27, the mean = 30.96 and the standard deviation = 10.58 given that $k$ = 10. Where $k$ is the number of simulations for each sample.

**Table 1.** Simulated data for the disease group.

| $c_1$ | $c_2$ | $c_3$ | $c_4$ | $c_5$ | $c_6$ | $c_7$ | $c_8$ | $c_9$ | $c_{10}$ |
|---|---|---|---|---|---|---|---|---|---|
| 46.1566 | 24.9042 | 46.9927 | 29.8675 | 25.2785 | 30.8228 | 22.7533 | 24.363 | 29.6894 | 13.9009 |
| 21.2799 | 34.1291 | 42.1188 | 31.4932 | 33.6757 | 18.2723 | 22.1952 | 49.3525 | 46.2031 | 49.671 |
| 47.1303 | 30.5768 | 30.518 | 45.4015 | 32.3025 | 23.9524 | 16.9939 | 40.0623 | 10.4986 | 34.2949 |
| 23.587 | 36.1635 | 31.943 | 32.0898 | 27.8432 | 10.0288 | 28.203 | 25.4455 | 36.9601 | 33.8601 |
| 32.5036 | 46.7256 | 28.132 | 29.1522 | 51.0953 | 30.5528 | 34.111 | 20.3037 | 41.3478 | 34.1312 |
| 39.3499 | 49.8787 | 22.8948 | 17.7508 | 33.7767 | 16.7141 | 47.7882 | 18.4173 | 24.499 | 24.9916 |
| 27.7979 | 27.2112 | 23.5745 | 22.0293 | 19.0153 | 37.9887 | 46.8897 | 32.489 | 41.2158 | 35.2746 |
| 31.4778 | 9.0139 | 23.6648 | 24.9898 | 31.9895 | 22.7037 | 32.1337 | 32.9346 | 27.7642 | 30.2984 |
| 26.1055 | 35.3076 | 33.2336 | 19.5483 | 32.9811 | 0.0334 | 24.751 | 68.7353 | 46.6003 | 14.3416 |
| 22.3024 | 32.34 | 26.1644 | 31.9692 | 33.0355 | 36.1969 | 18.5346 | 27.057 | 22.5413 | 34.9285 |
| 43.4464 | 30.52 | 18.1058 | 52.6543 | 35.5163 | 31.2813 | 20.1885 | 46.8498 | 15.0211 | 47.8509 |
| 22.2347 | 46.3877 | 19.7343 | 39.3903 | 19.5559 | 35.2009 | 25.0462 | 13.1057 | 35.0454 | 30.5387 |
| 47.6567 | 35.0258 | 23.2739 | 30.2782 | 32.4309 | 37.329 | 22.0514 | 38.6589 | 27.3124 | 25.0089 |
| 21.5468 | 26.4474 | 23.743 | 37.6406 | 33.1798 | 54.0146 | 37.2806 | 49.7954 | 16.1582 | 34.0969 |
| 48.2125 | 51.6615 | 30.1362 | 20.0116 | 29.284 | 48.7878 | 28.5845 | 34.9387 | 45.9698 | 29.2652 |
| 17.2076 | 50.856 | 14.453 | 18.2815 | 16.194 | 24.1627 | 28.8266 | 30.7418 | 41.5284 | 18.059 |
| 35.814 | 27.5266 | 20.5512 | 39.6783 | 31.3153 | 31.1173 | 36.2106 | 39.5762 | 46.0828 | 44.6502 |
| 34.1454 | 33.6547 | 33.7583 | 36.9679 | 23.2741 | 24.6134 | 27.161 | 37.796 | 17.8681 | 23.8681 |
| 27.8934 | 30.6502 | 22.5171 | 29.3292 | 32.4533 | 25.1776 | 29.9907 | 37.0937 | 36.4184 | 46.3087 |
| 30.988 | 35.2074 | 5.5657 | 51.0641 | 38.3618 | 26.2849 | 19.7737 | 26.8203 | 27.8187 | 7.8091 |
| 20.2051 | 29.3838 | 29.7345 | 37.2696 | 23.3611 | 17.3978 | 14.072 | 36.5702 | 33.102 | 32.488 |
| 39.5522 | 33.4252 | 37.0668 | 34.7706 | 27.7956 | 32.3243 | 30.6686 | 41.1511 | 21.3142 | 38.1397 |
| 35.3807 | 16.0406 | 40 | 36.8218 | 36.9183 | 31.8882 | 31.9571 | 24.1956 | 39.3912 | 27.1038 |
| 44.7964 | 22.1245 | 48.5002 | 42.257 | 19.9956 | 13.4868 | 50.2396 | 24.393 | 32.73 | 34.6324 |
| 28.535 | 58.7351 | 20.1339 | 44.5558 | 33.7764 | 26.7785 | 30.5453 | 39.9822 | 40.4087 | 37.8987 |
| 15.32 | 25.1028 | 32.6341 | 43.7344 | 28.3943 | 38.3414 | 11.665 | 20.4557 | 41.0132 | 27.5981 |
| 44.5328 | 40.3318 | 32.3106 | 25.3015 | 24.2536 | 20.2293 | 46.9336 | 33.98 | 31.2518 | 28.6054 |

**Table 2** is computed parameters from the 10 simulated data. Here the mean, standard deviation, 1st quartile ($Q_1$), median, 3rd quartile ($Q_3$) and the interquartile range (IQR) of each of the 10 simulated data were computed and the mean is the average of the data. *StDev* is the standard deviation. $Q_1$ is the 1st quarter or the 25th percentile of the data. Median is the middle value or the 2nd quartile or the 50th percentile of the data. $Q_3$ is the 3rd quartile or the 75th percentile of the data. IQR is the difference between the $Q_3$ and $Q_1$ *i.e.*, IQR = $Q_3 - Q_1$.

**Table 3** is random numbers generated from a normal distribution with parameters-sample size $n$ = 27, the mean = 30 and the standard deviation = 5.46 with $k$ = 10.

**Table 2.** Computed parameters from the disease group.

| Variable | Mean | StDev | $Q_1$ | Median | $Q_3$ | IQR |
|---|---|---|---|---|---|---|
| *x* | 32.41 | 10.25 | 22.30 | 31.48 | 43.45 | 21.14 |
| *h* | 34.05 | 11.23 | 27.21 | 33.43 | 40.33 | 13.12 |
| *k* | 28.20 | 9.70 | 22.52 | 28.13 | 33.23 | 10.72 |
| *l* ` | 33.49 | 9.70 | 25.30 | 32.09 | 39.68 | 14.38 |
| *o* | 29.89 | 7.24 | 24.25 | 31.99 | 33.68 | 9.42 |
| *p* | 27.62 | 11.40 | 20.23 | 26.78 | 35.20 | 14.97 |
| *r* | 29.09 | 10.24 | 22.05 | 28.58 | 34.11 | 12.06 |
| *w* | 33.90 | 11.74 | 24.39 | 33.98 | 39.98 | 15.59 |
| *v* | 32.44 | 10.47 | 24.50 | 33.10 | 41.22 | 16.72 |
| *z* | 31.10 | 10.14 | 25.01 | 32.49 | 35.27 | 10.27 |

**Table 3.** Simulated data for the normal group.

| $c_1$ | $c_2$ | $c_3$ | $c_4$ | $c_5$ | $c_6$ | $c_7$ | $c_8$ | $c_9$ | $c_{10}$ |
|---|---|---|---|---|---|---|---|---|---|
| 32.8692 | 30.7918 | 37.2667 | 16.7622 | 30.1366 | 41.7429 | 22.3522 | 40.3443 | 40.8101 | 29.0347 |
| 31.2128 | 35.0388 | 42.047 | 25.5956 | 23.9902 | 37.5143 | 27.3502 | 33.8674 | 26.7182 | 30.17 |
| 26.8548 | 19.9477 | 28.7859 | 36.2742 | 33.7429 | 35.2647 | 33.0093 | 33.8347 | 30.177 | 25.0604 |
| 28.3882 | 31.514 | 20.7621 | 24.3379 | 26.3542 | 32.3002 | 36.1576 | 24.9404 | 28.5576 | 39.2902 |
| 36.6662 | 30.297 | 38.2554 | 31.6718 | 32.3266 | 24.9395 | 36.9424 | 30.6826 | 29.9622 | 48.4579 |
| 26.6789 | 24.8178 | 32.3347 | 29.5581 | 30.8839 | 34.3397 | 33.6549 | 35.1657 | 31.1464 | 28.1627 |
| 21.545 | 31.0153 | 34.2887 | 27.2154 | 38.5984 | 24.7763 | 27.3764 | 24.1752 | 40.6695 | 42.4859 |
| 24.1657 | 27.1545 | 34.7675 | 30.6038 | 22.5373 | 33.1506 | 21.4981 | 22.4504 | 40.6367 | 27.2556 |
| 28.5745 | 34.6841 | 39.1893 | 27.9174 | 32.9548 | 29.8816 | 28.3716 | 34.9552 | 20.3387 | 29.917 |
| 34.1356 | 27.1955 | 35.3834 | 32.0297 | 36.2353 | 23.8584 | 22.5997 | 35.8457 | 26.1652 | 28.1878 |
| 30.6207 | 36.3776 | 22.2232 | 37.488 | 24.1201 | 31.908 | 22.647 | 24.0633 | 31.8018 | 39.9472 |
| 28.1189 | 43.1577 | 28.1693 | 22.5995 | 23.482 | 18.6995 | 42.8021 | 40.0288 | 35.2402 | 38.9854 |
| 25.6035 | 31.0938 | 23.8768 | 20.4452 | 36.7756 | 25.4246 | 21.3635 | 28.1414 | 35.3096 | 29.555 |
| 21.2667 | 43.2111 | 22.849 | 41.8452 | 35.296 | 30.3376 | 40.7888 | 24.0153 | 18.9364 | 24.2351 |
| 24.0752 | 28.0961 | 26.8861 | 27.4234 | 32.2276 | 23.5675 | 24.8017 | 30.1957 | 29.3453 | 34.6046 |
| 35.9706 | 30.977 | 21.9998 | 24.034 | 30.5767 | 25.7669 | 35.9832 | 37.6228 | 22.4463 | 30.0228 |
| 30.8747 | 16.9055 | 28.5309 | 26.1719 | 22.9281 | 39.2016 | 21.9753 | 30.829 | 33.1866 | 34.1112 |
| 27.7715 | 35.1348 | 32.7291 | 36.4564 | 30.2984 | 41.7592 | 35.1538 | 33.751 | 34.8372 | 34.1248 |
| 28.6439 | 31.1899 | 27.4617 | 36.4554 | 26.8265 | 24.9297 | 25.6044 | 20.2664 | 21.662 | 38.6479 |
| 24.5744 | 32.6061 | 17.7682 | 33.5074 | 42.5921 | 28.2345 | 34.5526 | 29.9673 | 31.0932 | 22.2289 |
| 38.0338 | 34.2826 | 30.6231 | 30.2866 | 33.7578 | 34.7125 | 20.2292 | 27.3069 | 35.8389 | 23.4028 |
| 34.2659 | 44.6651 | 16.1165 | 31.541 | 24.5278 | 35.8873 | 23.3711 | 20.5715 | 33.6894 | 28.5468 |
| 30.753 | 28.6222 | 37.3659 | 25.6089 | 24.0786 | 33.2956 | 21.4613 | 31.2423 | 24.4339 | 28.9205 |
| 37.6636 | 36.7431 | 37.1959 | 23.0103 | 28.9785 | 28.5097 | 28.8184 | 34.9921 | 33.7258 | 33.6018 |
| 41.7951 | 23.519 | 37.6545 | 25.959 | 24.1875 | 27.8784 | 32.629 | 30.366 | 28.4286 | 37.4612 |
| 33.925 | 28.6911 | 29.451 | 27.8078 | 33.4418 | 19.7654 | 34.997 | 28.4154 | 41.6175 | 21.2209 |
| 24.6987 | 28.8511 | 20.9158 | 28.4334 | 33.0239 | 36.3741 | 32.0921 | 28.7408 | 28.0413 | 40.8146 |

**Table 4** is computed parameters from the normal group samples.

**Table 5** contains calculated Modified *Z*-Statistic, COPA and OS from the first sample of the disease group. From the table, $Y_1$ are the values of the first simulated data from the sample. $Q_1$ of $Y_1$ = 22.3 is the first quartile of $Y_1$, $Med_1$ = 31.5 is the median and 2nd quarter of $Y_1$. $Q_3$ of $Y_1$ = 43.45 is the third quartile of $Y_1$ and IQR of $Y_1$ = 21.14 is the interquartile range of $Y_1$.

**Table 4.** Computed parameters from the normal group samples.

| Variable | Mean | StDev | $Q_1$ | Median | $Q_3$ | IQR |
|---|---|---|---|---|---|---|
| *c*1 | 29.99 | 5.31 | 25.60 | 28.64 | 34.14 | 8.53 |
| *c*2 | 31.35 | 6.42 | 28.10 | 31.02 | 35.04 | 6.94 |
| *c*3 | 29.81 | 7.19 | 22.85 | 29.45 | 37.20 | 14.35 |
| *c*4 | 28.93 | 5.70 | 25.60 | 27.92 | 32.03 | 6.43 |
| *c*5 | 30.18 | 5.39 | 24.19 | 30.58 | 33.74 | 9.56 |
| *c*6 | 30.52 | 6.27 | 24.94 | 30.34 | 35.26 | 10.33 |
| *c*7 | 29.21 | 6.64 | 22.60 | 28.37 | 35.00 | 12.40 |
| *c*8 | 30.25 | 5.57 | 24.94 | 30.37 | 34.96 | 10.01 |
| *c*9 | 30.92 | 6.26 | 26.72 | 31.09 | 35.24 | 8.52 |
| *c*10 | 32.16 | 6.82 | 28.16 | 30.02 | 38.65 | 10.49 |

**Table 5.** Calculated Modified *Z*, COPA and OS.

| **$Y_1$** | Med 1 | $Q_1$ of $Y_1$ | $Q_3$ of $Y_1$ | IQR of $Y_1$ | $Y_1$-*Med*1 | $Y_1$-$Med_1$ | $Mad_1$ | ($Y_1$-$Med_1$/ $Mad_1$) = *M*' | **COPA1** | Constant 1 | *M*' × Constant 1 | **Z-Value 1** | Outliers 1 | *M*'1 of $O_1$ | **OS1** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 46.1566 | 31.48 | 22.3 | 43.45 | 21.14 | 14.6766 | 14.6766 | 9.18 | 1.59876 | −1.7603 | 0.6745 | 1.07836 | 0 | 46.1566 | 1.59876 | 1.39 |
| 21.2799 | 31.48 | | | | −10.2 | 10.2001 | 9.18 | −1.11112 | | 0.6745 | 0.74945 | | 47.1303 | −1.1111 | |
| 47.1303 | 31.48 | | | | 15.6503 | 15.6503 | 9.18 | 1.70483 | | 0.6745 | 1.1499 | | 47.6567 | 1.76217 | |
| 23.587 | 31.48 | | | | −7.893 | 7.893 | 9.18 | −0.8598 | | 0.6745 | 0.57994 | | 48.2125 | −0.8598 | |
| 32.5036 | 31.48 | | | | 1.0236 | 1.0236 | 9.18 | 0.1115 | | 0.6745 | 0.07521 | | | | |
| 39.3499 | 31.48 | | | | 7.8699 | 7.8699 | 9.18 | 0.85729 | | 0.6745 | 0.57824 | | | | |
| 27.7979 | 31.48 | | | | −3.6821 | 3.6821 | 9.18 | −0.4011 | | 0.6745 | 0.27054 | | | | |
| 31.4778 | 31.48 | | | | −0.0022 | 0.0022 | 9.18 | −0.00024 | | 0.6745 | 0.00016 | | | | |
| 26.1055 | 31.48 | | | | −5.3745 | 5.3745 | 9.18 | −0.58546 | | 0.6745 | 0.39489 | | | | |
| 22.3024 | 31.48 | | | | −9.1776 | 9.1776 | 9.18 | −0.99974 | | 0.6745 | 0.67432 | | | | |
| 43.4464 | 31.48 | | | | 11.9664 | 11.9664 | 9.18 | 1.30353 | | 0.6745 | 0.87923 | | | | |
| 22.2347 | 31.48 | | | | −9.2453 | 9.2453 | 9.18 | −1.00711 | | 0.6745 | 0.6793 | | | | |
| 47.6567 | 31.48 | | | | 16.1767 | 16.1767 | 9.18 | 1.76217 | | 0.6745 | 1.18858 | | | | |
| 21.5468 | 31.48 | | | | −9.9332 | 9.9332 | 9.18 | −1.08205 | | 0.6745 | 0.72984 | | | | |
| 48.2125 | 31.48 | | | | 16.7325 | 16.7325 | 9.18 | 1.82271 | | 0.6745 | 1.22942 | | | | |
| 17.2076 | 31.48 | | | | −14.272 | 14.2724 | 9.18 | −1.55473 | | 0.6745 | 1.04866 | | | | |
| 35.814 | 31.48 | | | | 4.334 | 4.334 | 9.18 | 0.47211 | | 0.6745 | 0.31844 | | | | |
| 34.1454 | 31.48 | | | | 2.6654 | 2.6654 | 9.18 | 0.29035 | | 0.6745 | 0.19584 | | | | |
| 27.8934 | 31.48 | | | | −3.5866 | 3.5866 | 9.18 | −0.3907 | | 0.6745 | 0.26353 | | | | |
| 30.988 | 31.48 | | | | −0.492 | 0.492 | 9.18 | −0.05359 | | 0.6745 | 0.03615 | | | | |
| 20.2051 | 31.48 | | | | −11.275 | 11.2749 | 9.18 | −1.2282 | | 0.6745 | 0.82842 | | | | |
| 39.5522 | 31.48 | | | | 8.0722 | 8.0722 | 9.18 | 0.87932 | | 0.6745 | 0.5931 | | | | |
| 35.3807 | 31.48 | | | | 3.9007 | 3.9007 | 9.18 | 0.42491 | | 0.6745 | 0.2866 | | | | |
| 44.7964 | 31.48 | | | | 13.3164 | 13.3164 | 9.18 | 1.45059 | | 0.6745 | 0.97842 | | | | |
| 28.535 | 31.48 | | | | −2.945 | 2.945 | 9.18 | −0.32081 | | 0.6745 | 0.21638 | | | | |
| 15.32 | 31.48 | | | | −16.16 | 16.16 | 9.18 | −1.76035 | | 0.6745 | 1.18736 | | | | |
| 44.5328 | 31.48 | | | | 13.0528 | 13.0528 | 9.18 | 1.42187 | | | 0.95905 | | | | |

$(Y_1 - Med_1)$ is the deviation of each of the simulated value of $Y_1$ from the median of $Y_1$.

$|Y_1 - Med_1|$ is the absolute of the values calculated in $(Y_1 - Med_1)$. $Mad_1 = 9.18$ is the median of the absolute values in $|Y_1 - Med_1|$. Each value in $Y_1$ is first standardized by centering and putting all the values on the same scale which facilitates comparison across the values *i.e.* $\frac{Y_1 - Med_1}{Mad_1} = M'$ is the centered value for comparison. The choice of COPA is based on the $M'$ corresponding to any value in $Y_1$ that is less than $n_1 = 27$ the total number of values in $Y_1$. Here, COPA value is −1.7603 corresponding to 15.32 in $Y_1$ which is less than 27. The use of the constant (75) the $r^{th}$ percentile to multiply $M'$ was ignored since it will not affect the order of the values.

The $Z$-Value = 0 is calculated by multiplying $M'$ by a constant 0.6745 and the choice of $Z$ is based on any of $0.6745 \times M'$ that is greater than 3.5 in absolute value. *i.e.*

$$Z = \text{any of } \left| \frac{Y_1 - Med_1}{Mad_1} \cdot 0.6745 \right| > 3.5.$$

OS = 1.4 is calculated by summing up all the $M'$ corresponding to the outlier set in $Y_1$.

### 3.1. Two Sample *t*-test and Confidence Interval

Here, a two sample *t*-test was conducted to obtain the *t*-values, confidence interval and its corresponding *P*-values using the sample means and standard deviations from both the disease and normal group samples of the simulated data assuming equal variances.

**Table 6** is the parameters from both the normal group and the disease group data used for the two-sample *t*-Test.

The Test Difference = mu (1) − mu (2).

Estimate for difference: 4.00000.

95% CI for difference: (−0.31762; 8.31762).

*t*-test of difference = 0 (vs not = 0): *t*-value = 1.86, *P*-value = 0.069 df = 52.

Both use Pooled StDev = 7.9057.

From the *t*-test above, sample 1 is the first sample from the disease group with parameters: $n$ = 27, mean = 32, standard deviation = 10 and squared error = 1.9. Sample 2 also is the first sample from the normal group with parameters: $n$ = 27, mean =28, standard deviation = 5 and squared error = 0.96, mu (1) − mu (2) is the difference between the mean of sample 1 and the mean of sample 2 and this is equal to 4. From the test also, we are 95% confident that the mean will lie between the interval (−0.31762; 8.31762).

**The test hypothesis is**:

$H_o$: mu(1) = mu(2).

vs

$H_1$: mu(1) ≠ mu(2).

*t*-value = 1.86, *P*-value = 0.069. The degree of freedom *df* is calculated by adding the two sample sizes and subtracting 2 from it. *i.e.* $(n_1 + n_2) - 2$.

Therefore *df* = ((27 +27) – 2) = 52 with a pooled standard deviation of 7.9057 which is the square root of the pooled sample variances of the two samples assuming equal variances *i.e.* homogeneity of variances and since the *P*-value = 0.069 is greater than the significant level $\alpha$ = 0.05, hence $H_o$ is rejected which implies that the mean ≠ 0.

**Table 7** shows calculated ORT and TORT. ORT and TORT utilizes information from both the normal and the disease group sample.

From **Table 16**, $Y_1$ represents the values of the first simulated data from the first sample regarded as the disease group. $Q_1$ = 22.3 is the first quartile of $Y_1$, $Medd_1 = 31.5$ is the median of the disease group $Y_1$. $Q_3$ = 43.45 is

**Table 6.** Two-sample *t*-test and confidence interval for sample 1.

| Sample | *N* | Mean | StDev | SE Mean |
|---|---|---|---|---|
| 1 | 27 | 32.0 | 10.0 | 1.9 |
| 2 | 27 | 28.00 | 5.00 | 0.96 |

**Table 7.** Calculated ORT and TORT.

| $Y_1$ | $Q_1$ of $Y_1$ | $Q_3$ of $Y_1$ | IQR of $Y_1$ | *Medd*1 | *Madd*1 | *Outliers*1 | *Medo*1 | \|*Outliers*1-*Medo*1\| | *Mado*1 | $X_1$ | *Medc*1 | *Madc*1 | *Madd*1 × *Madc*1 | *Madc*1 × *Mado*1 | *Outliers*1 − *Medc*1 | *Outliers*1 − *Medc*1/ *Madc*1 × *Madd*1 | *Outliers*1 − *Medc*1/ *Madc*1 × *Mado*1 | ORT1 | TORT1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 46.1566 | 22.3 | 43.5 | 21.1 | 31.48 | 9.18 | 46.157 | 47.394 | 1.2369 | 0.541 | 33.9 | 26.47 | 3.939 | 36.1554 | 2.13112 | 19.6866 | 0.544499 | 9.2377 | 2.30328 | 39.0762 |
| 21.2799 | | | | | | 47.13 | 47.394 | 0.2632 | | 22.2 | 26.47 | | 36.1554 | 2.13112 | 20.6603 | 0.57143 | 9.6946 | | |
| 47.1303 | | | | | | 47.657 | 47.394 | 0.2632 | | 26.5 | 26.47 | | 36.1554 | 2.13112 | 21.1867 | 0.585989 | 9.9416 | | |
| 23.587 | | | | | | 48.213 | 47.394 | 0.819 | | 30.4 | 26.47 | | 36.1554 | 2.13112 | 21.7425 | 0.601362 | 10.2024 | | |
| 32.5036 | | | | | | | | | | 25.9 | | | | | | | | | |
| 39.3499 | | | | | | | | | | 33.2 | | | | | | | | | |
| 27.7979 | | | | | | | | | | 30.3 | | | | | | | | | |
| 31.4778 | | | | | | | | | | 35.8 | | | | | | | | | |
| 26.1055 | | | | | | | | | | 31.4 | | | | | | | | | |
| 22.3024 | | | | | | | | | | 22.9 | | | | | | | | | |
| 43.4464 | | | | | | | | | | 26.6 | | | | | | | | | |
| 22.2347 | | | | | | | | | | 30.8 | | | | | | | | | |
| 47.6567 | | | | | | | | | | 28.1 | | | | | | | | | |
| 21.5468 | | | | | | | | | | 34.8 | | | | | | | | | |
| 48.2125 | | | | | | | | | | 24.3 | | | | | | | | | |
| 17.2076 | | | | | | | | | | 32.9 | | | | | | | | | |
| 35.814 | | | | | | | | | | 30.7 | | | | | | | | | |
| 34.1454 | | | | | | | | | | 26.3 | | | | | | | | | |
| 27.8934 | | | | | | | | | | 21.8 | | | | | | | | | |
| 30.988 | | | | | | | | | | 31.5 | | | | | | | | | |
| 20.2051 | | | | | | | | | | 25.8 | | | | | | | | | |
| 39.5522 | | | | | | | | | | 26.1 | | | | | | | | | |
| 35.3807 | | | | | | | | | | 22.8 | | | | | | | | | |
| 44.7964 | | | | | | | | | | 19.3 | | | | | | | | | |
| 28.535 | | | | | | | | | | 26.3 | | | | | | | | | |
| 15.32 | | | | | | | | | | 21.6 | | | | | | | | | |
| 44.5328 | | | | | | | | | | 25.8 | | | | | | | | | |

the third quartile of the disease group and IQR = 21.14 is the interquartile range of the disease group $Y_1$ and $Madd1 = 9.18$ is the median absolute deviation of the disease group. Outliers1 are the outliers from the disease group. ORT concentrates on only the outlier set $O_i$ from the disease group. $Medo1 = 47.39$ is the median of the outlier set $O_i$ from the disease group.

In order to standardize and put the outliers on the same scale for comparison across the outliers, $|Outliers1 - Medo1|$ is computed which is the deviation of each outlier point from the median of the outlier set $O_i$ in absolute value. $Mado1 = 0.5411$ is the median absolute deviation of the outlier set which is calculated as: $Mado1 = median|Outliers - Med_o|$.

$X_1$ are the values of the first simulated data from the second sample regarded as the normal or control group. $Medc1 = 26.5$ is the median of the control group $X_1$ and $Mado1 = 3.39$ is the median absolute deviation of the control group. $(Madd1 \times Madc1) = 36.1554$ is the product of the median absolute deviation of the disease group and the median absolute deviation of the control group. $Madc1 \times Mado1 = 2.1311$ is the product of the median absolute deviation of the control group and the median absolute deviation of the outlier set $O_i$ $(Outliers1 - Medc1)$ is the deviation of the outliers from the median of the control group.

These values were calculated to facilitate comparison across the two samples.

The value for ORT= 2.303 is obtained by dividing $(Outliers1 - Medc1)$ by $(Madd1 \times Madc1)$ while the value for TORT = 39.08 is obtained by diving $(Outliers1 - Medc1)$ by $Madc1 \times Mado1$.

**Table 8** shows the summary of the computed values of all the simulated data by the six outlier methods. From the table, we have 80 samples of the simulated data from the disease group. The value for each of the outlier method *Z*, *t*-distribution, COPA, OS, ORT and TORT was calculated for each sample for comparison among the outlier methods.

**Table 9** shows the computed *P*-values. Here, *P*-values were computed for all the values computed by the outlier methods. The *P*-values were generated from the standard normal *Z*-distribution and *t*-distribution. This is based on the assumption that the Modified *Z*-Statistic, COPA and OS are assumed to follow a normal distribution while the *t*-Statistic, ORT and TORT are assumed to follow a *t*-distribution.

**Table 8.** Summary of computed values for the various outlier techniques.

| SAMPLES | Z-VALUE | *t*-VALUE | COPA | OS | ORT | TORT |
|---|---|---|---|---|---|---|
| $Y_1$ | 0 | 1.86 | −1.76034 | 1.39 | 2.30328 | 39.076 |
| $Y_2$ | 0 | 2.49 | −3.92618 | 12.7434 | 5.55294 | 10.952 |
| $Y_3$ | 0 | 0.43 | −3.27801 | 0.8372 | 0.60208 | 3.157 |
| $Y_4$ | 0 | 3.72 | −2.01955 | 5.5686 | 0.2825 | 2.523 |
| $Y_5$ | 0 | 2.25 | −4.39303 | 5.3134 | 1.41896 | 0 |
| $Y_6$ | 0 | 0.43 | −4.08301 | 0.8769 | 0.15131 | 0.045 |
| $Y_7$ | 0 | 0.89 | −2.64926 | 9.4938 | 2.73361 | 20.423 |
| $Y_8$ | 0 | 2 | −2.91553 | 11.0079 | 4.33481 | 21.073 |
| $Y_9$ | 0 | 1.99 | −2.78486 | 2.0928 | 1.39099 | 43.629 |
| $Y_{10}$ | 0 | 1.86 | −4.58225 | 4.025 | 1.96214 | 6.286 |
| $Y_{11}$ | 0 | −0.52 | −2.97241 | 2.7904 | 0.71192 | 4.745 |
| $Y_{12}$ | 0 | 1.13 | −3.10803 | 4.0509 | 1.2525 | 0 |
| $Y_{13}$ | 0 | 0 | −3.4866 | 3.0625 | 0.78095 | 0.213 |
| $Y_{14}$ | 0 | −0.43 | −2.55996 | 0.0082 | −0.26098 | −0.099 |
| $Y_{15}$ | 0 | −1.04 | −2.46655 | 1.2416 | 0.11346 | 0.043 |
| $Y_{16}$ | 0 | 1.1 | −2.17537 | −3.2457 | −0.77117 | −1.65 |
| $Y_{17}$ | 0 | −0.45 | −2.15159 | 3.3852 | 0.39417 | 0.628 |
| $Y_{18}$ | 0 | 1.04 | −3.28757 | 4.8094 | 2.51337 | 5.166 |
| $Y_{19}$ | 0 | 0.96 | −2.86551 | −6.1141 | −0.95734 | −0.87 |
| $Y_{20}$ | 0 | 0.67 | −2.74892 | −4.727 | −1.27434 | −2.597 |
| $Y_{21}$ | 0 | 1.1 | −1.74513 | 3.2357 | 0.93583 | 0 |
| $Y_{22}$ | 0 | 1.13 | −3.85211 | 6.5117 | 2.32341 | 11.295 |
| $Y_{23}$ | 0 | 1.13 | −2.01992 | −1.8813 | −0.14041 | −0.561 |
| $Y_{24}$ | 0 | 0 | −3.08141 | 0.7643 | 0.2648 | −0.223 |
| $Y_{25}$ | 0 | −0.6 | −1.67968 | 2.8426 | 0.49957 | 0 |
| $Y_{26}$ | 0 | 1.13 | −1.37521 | 4.6342 | 1.04703 | 632.674 |
| $Y_{27}$ | 0 | 0.98 | −1.06263 | 4.4619 | 0.7465 | 4.463 |
| $Y_{28}$ | 0 | 1.69 | −1.72721 | 4.2818 | 1.61369 | 46.904 |
| $Y_{29}$ | 0 | 2.82 | −1.0534 | 9.4872 | 3.49909 | 167.496 |
| $Y_{30}$ | 0 | 2.1 | −1.26181 | 9.0953 | 9.18679 | 5.497 |
| $Y_{31}$ | 0 | 0.89 | −1 | 3.8121 | 1.09248 | 0 |
| $Y_{32}$ | 0 | −0.98 | −4.48556 | −8.0926 | −1.80663 | −4.113 |
| $Y_{33}$ | 0 | 0.5 | −1.2577 | −2.8623 | −0.73876 | 0 |
| $Y_{34}$ | 0 | −2.4 | −2.06076 | −3.9971 | −2.42707 | −22.154 |
| $Y_{35}$ | 0 | −2.08 | −2.70112 | −7.4548 | −2.9424 | −56.16 |

**Continued**

| | | | | | | |
|---|---|---|---|---|---|---|
| $Y_{36}$ | 0 | −1.22 | −1.24659 | 0 | 0 | 0 |
| $Y_{37}$ | 0 | −1.69 | −1.25708 | 0 | 0 | 0 |
| $Y_{38}$ | 0 | 0.6 | −1.93532 | 0 | 0 | 0 |
| $Y_{39}$ | 0 | −2.93 | −3.52858 | −9.1664 | −3.43498 | −3.257 |
| $Y_{40}$ | 0 | −2 | −0.91637 | −2.8195 | −0.82398 | 0 |
| $Y_{41}$ | 0 | 0.41 | −3.36643 | −2.0476 | −2.86471 | −17.903 |
| $Y_{42}$ | 0 | 0.56 | −3.4781 | −3.4781 | −1.64245 | 0 |
| $Y_{43}$ | 0 | 0.49 | −5.07365 | −5.0737 | −1.14229 | 0 |
| $Y_{44}$ | 0 | 2.44 | 0.8284 | 4.1342 | 0.71059 | 119.372 |
| $Y_{45}$ | 0 | 1.56 | 1.56 | −8.3637 | −1.80281 | −166.776 |
| $Y_{46}$ | 0 | 0.96 | −1.55075 | 0.4615 | 0.57185 | 0.198 |
| $Y_{47}$ | 0 | 4.41 | −2.49374 | 0 | 0 | 0 |
| $Y_{48}$ | 0 | 2.82 | −2.63485 | 0 | 0 | 0 |
| $Y_{49}$ | 0 | 0.46 | 0.46 | −5.2878 | −1.41388 | −33.607 |
| $Y_{50}$ | 0 | 0.46 | −2.5987 | −2.5987 | −0.56607 | 0 |
| $Y_{51}$ | 0 | −2.39 | −2.10795 | −7.8008 | −3.68414 | −21.252 |
| $Y_{52}$ | 0 | −2.44 | −1.66542 | −2.3363 | −0.44473 | 0 |
| $Y_{53}$ | 0 | −0.98 | −0.50455 | −5.1537 | −1.15351 | −24.42 |
| $Y_{54}$ | 0 | 1.47 | −2.52821 | −11.2182 | −2.54795 | −2.425 |
| $Y_{55}$ | 0 | 0 | −2.70932 | −14.1094 | −1.79082 | −4.729 |
| $Y_{56}$ | 0 | −0.43 | −2.17732 | −10.2926 | −2.1721 | −7.973 |
| $Y_{57}$ | 0 | 0.49 | −1.09271 | −2.754 | −1.03836 | −11.2 |
| $Y_{58}$ | 0 | 0.91 | −3.42848 | −9.9813 | −2.53859 | −6.818 |
| $Y_{59}$ | 0 | −1.82 | −1.90071 | −9.737 | −2.27291 | −7.208 |
| $Y_{60}$ | 0 | −0.91 | −1.14792 | −7.0358 | −1.90352 | −3.406 |
| $Y_{61}$ | 0 | 3.49 | −0.11938 | 0.0865 | 1.08222 | 1.604 |
| $Y_{62}$ | 0 | 2.73 | −1 | −4.3347 | 0.35235 | 0.515 |
| $Y_{63}$ | 0 | 0.48 | −1.16752 | −6.618 | −1.47642 | −5.356 |
| $Y_{64}$ | 0 | 0 | −1.75915 | 9.3197 | 1.79452 | 0 |
| $Y_{65}$ | 0 | 0.38 | −1.34253 | 8.3705 | 2.23081 | 23.572 |
| $Y_{66}$ | 0 | 2.19 | −1.71693 | 7.8552 | 0.11634 | 0.145 |
| $Y_{67}$ | 0 | 3.36 | −3.68369 | −3.3687 | 0.22073 | 0.489 |
| $Y_{68}$ | 0 | 2.88 | −0.84967 | −2.4492 | −0.42245 | 0 |
| $Y_{69}$ | 0 | −0.85 | −1.41307 | −6.2249 | −1.54908 | −10.655 |
| $Y_{70}$ | 0 | 2.98 | −1.86298 | −2.0592 | −0.02425 | −0.005 |
| $Y_{71}$ | 0 | 0 | −1.44536 | 3.4758 | 0.47375 | 0 |
| $Y_{72}$ | 0 | 0 | −1.44536 | 3.4758 | 0.47375 | 0 |
| $Y_{73}$ | 0 | −0.46 | −0.52618 | −5.313 | −1.01075 | −6.144 |
| $Y_{74}$ | 0 | −0.8 | −4.23945 | −11.4121 | −1.23036 | −0.389 |
| $Y_{75}$ | 0 | −0.43 | −2.88634 | −2.8863 | −0.47865 | 0 |
| $Y_{76}$ | 0 | 1.84 | 1.08439 | 1.0844 | 0.61602 | 0.741 |
| $Y_{77}$ | 0 | −0.37 | −2.32299 | −9.2371 | −2.65121 | −4.663 |
| $Y_{78}$ | 0 | −0.49 | −0.14608 | 0 | 0 | 0 |
| $Y_{79}$ | 0 | 0.86 | −1.01972 | −2.8507 | −0.3255 | 0 |
| $Y_{80}$ | 0 | 0.43 | −2.21624 | −2.2162 | −0.3952 | 0 |

**Table 9.** Summary of computed *P*-values for the various outlier techniques.

| *Z* | *T* | COPA | OS | ORT | TORT |
|---|---|---|---|---|---|
| 1 | 0.069 | 0.0784 | 0.1645 | 0.0295 | 0.0001 |
| 1 | 0.016 | 0.0001 | 0.0001 | 0.0001 | 0.0001 |
| 1 | 0.672 | 0.001 | 0.4025 | 0.5523 | 0.004 |
| 1 | 0 | 0.0434 | 0.0001 | 0.7798 | 0.0181 |
| 1 | 0.028 | 0.0001 | 0.0001 | 0.1678 | 1 |
| 1 | 0.669 | 0.0001 | 0.3805 | 0.8809 | 0.9645 |
| 1 | 0.377 | 0.0081 | 0.0001 | 0.111 | 0.0001 |
| 1 | 0.051 | 0.0036 | 0.0001 | 0.0002 | 0.0001 |
| 1 | 0.052 | 0.0054 | 0.0364 | 0.176 | 0.0001 |
| 1 | 0.069 | 0.0001 | 0.0001 | 0.0605 | 0.0001 |
| 1 | 0.606 | 0.003 | 0.0053 | 0.4829 | 0.0001 |
| 1 | 0.265 | 0.0019 | 0.0001 | 0.2215 | 1 |
| 1 | 1 | 0.0005 | 0.0022 | 0.4419 | 0.833 |
| 1 | 0.672 | 0.0052 | 0.9935 | 0.7962 | 0.9219 |
| 1 | 0.304 | 0.0136 | 0.2144 | 0.9105 | 0.966 |
| 1 | 0.276 | 0.00031 | 0.0012 | 0.4476 | 0.111 |
| 1 | 0.658 | 0.0314 | 0.0007 | 0.6967 | 0.5355 |
| 1 | 0.304 | 0.001 | 0.0001 | 0.0185 | 0.0001 |
| 1 | 0.341 | 0.0042 | 0.0001 | 0.3472 | 0.3923 |
| 1 | 0.67 | 0.006 | 0.0001 | 0.2138 | 0.0153 |
| 1 | 0.276 | 0.081 | 0.0012 | 0.358 | 1 |
| 1 | 0.265 | 0.0001 | 0.0001 | 0.0282 | 0.0001 |
| 1 | 0.265 | 0.0434 | 0.0599 | 0.8594 | 0.5796 |
| 1 | 1 | 0.0021 | 0.4447 | 0.7933 | 0.8253 |
| 1 | 0.548 | 0.093 | 0.0045 | 0.6216 | 1 |
| 1 | 0.265 | 0.1691 | 0.0001 | 0.3047 | 0.0001 |
| 1 | 0.333 | 0.2879 | 0.0001 | 0.4621 | 0.0001 |
| 1 | 0.097 | 0.0841 | 0.0001 | 0.1187 | 0.0001 |
| 1 | 0.007 | 0.2922 | 0.0001 | 0.0017 | 0.0001 |
| 1 | 0.041 | 0.207 | 0.0001 | 0.0001 | 0.0001 |
| 1 | 0.377 | 0.2922 | 0.0001 | 0.2846 | 1 |
| 1 | 0.333 | 1 | 0.0001 | 0.0824 | 0.0003 |
| 1 | 0.616 | 0.2085 | 0.0042 | 0.4667 | 1 |
| 1 | 0.02 | 0.0393 | 0.0001 | 0.0225 | 0.0001 |
| 1 | 0.043 | 0.0069 | 0.0001 | 0.0068 | 0.0001 |
| 1 | 0.226 | 0.2125 | 1 | 1 | 1 |
| 1 | 0.097 | 0.287 | 1 | 1 | 1 |
| 1 | 0.548 | 0.053 | 1 | 1 | 1 |

**Continued**

| | | | | | |
|---|---|---|---|---|---|
| 1 | 0.005 | 0.0004 | 0.0001 | 0.002 | 0.0031 |
| 1 | 0.051 | 0.3595 | 0.0048 | 0.4174 | 1 |
| 1 | 0.68 | 0.0008 | 0.0406 | 0.0082 | 0.0001 |
| 1 | 0.575 | 0.0005 | 0.0005 | 0.1125 | 1 |
| 1 | 0.627 | 0.0001 | 0.0001 | 0.2637 | 1 |
| 1 | 0.018 | 0.4074 | 0.0001 | 0.4837 | 0.0001 |
| 1 | 0.125 | 0.1188 | 0.0001 | 0.083 | 0.0001 |
| 1 | 0.341 | 0.121 | 0.6444 | 0.5723 | 0.8446 |
| 1 | 0 | 0.0126 | 1 | 1 | 1 |
| 1 | 0.007 | 0.0084 | 1 | 1 | 1 |
| 1 | 0.65 | 0.6455 | 0.0001 | 0.1693 | 0.0001 |
| 1 | 0.65 | 0.0094 | 0.0094 | 0.5762 | 1 |
| 1 | 0.02 | 0.035 | 0.0001 | 0.0011 | 1 |
| 1 | 0.018 | 0.0958 | 0.0195 | 0.6602 | 1 |
| 1 | 0.333 | 0.6139 | 0.0001 | 0.2592 | 0.0001 |
| 1 | 0.149 | 0.0115 | 0.0001 | 0.0171 | 0.022 |
| 1 | 1 | 0.0067 | 0.0001 | 0.055 | 0.0001 |
| 1 | 0.668 | 0.0293 | 0.0001 | 0.0392 | 0.0001 |
| 1 | 0.627 | 0.2745 | 0.0059 | 0.3087 | 0.0001 |
| 1 | 0.366 | 0.0006 | 0.0001 | 0.0175 | 0.0001 |
| 1 | 0.074 | 0.0573 | 0.0001 | 0.0315 | 0.0001 |
| 1 | 0.366 | 0.251 | 0.0001 | 0.0681 | 0.0022 |
| 1 | 0.001 | 0.905 | 0.9311 | 0.2891 | 0.1208 |
| 1 | 0.009 | 0.2922 | 0.0001 | 0.7274 | 0.6109 |
| 1 | 0.633 | 0.243 | 0.0001 | 0.1518 | 0.0001 |
| 1 | 1 | 0.0786 | 0.0001 | 0.0844 | 1 |
| 1 | 0.704 | 0.1794 | 0.0001 | 0.0345 | 0.0001 |
| 1 | 0.033 | 0.086 | 0.0001 | 0.9083 | 0.8856 |
| 1 | 0.001 | 0.0002 | 0.0008 | 0.827 | 0.5237 |
| 1 | 0.006 | 0.3955 | 0.0143 | 0.6762 | 1 |
| 1 | 0.398 | 0.1576 | 0.0001 | 0.1335 | 0.0001 |
| 1 | 0.004 | 0.0625 | 0.0395 | 0.9808 | 0.9681 |
| 1 | 1 | 0.1484 | 0.0005 | 0.6396 | 1 |
| 1 | 1 | 0.1484 | 0.0005 | 0.6396 | 1 |
| 1 | 0.648 | 0.5988 | 0.0001 | 0.3215 | 0.0001 |
| 1 | 0.429 | 0.0001 | 0.0001 | 0.2296 | 0.8345 |
| 1 | 0.668 | 0.0039 | 0.0039 | 0.6362 | 1 |
| 1 | 0.072 | 0.2782 | 0.2782 | 0.5432 | 0.8235 |
| 1 | 0.716 | 0.0202 | 0.0001 | 0.0135 | 0.0004 |
| 1 | 0.627 | 0.8839 | 1 | 1 | 1 |
| 1 | 0.392 | 0.3079 | 0.0044 | 0.7474 | 1 |
| 1 | 0.668 | 0.0267 | 0.0267 | 0.6959 | 1 |

From **Table 18**, we observed that all the outlier methods have equal maximum *P*-value of 1 (one). Modified *Z* has a minimum *P*-value of 1 (one), *t* and COPA have a minimum *P*-value of 0 (Zero) while OS, ORT and TORT have the least minimum *P*-value of 0.0001. Modified *Z* has 80 true positives and 0 false positives, *t*-Statistic has 61 true positives and 19 False Positives, COPA has 39 true positives and 41 false positives, OS has 16 true positives and 64 false positives, ORT has 62 true positives and 18 false positives while TORT has 42 true positives and 38 false positives. The Mean of the *P*-values of the different outlier methods were computed giving the following results: $Z = 1$, t = 0.360188, COPA = 0.144847, OS = 0.134311, ORT = 0.388911 and TORT = 0.472614. From these values, we can see that OS has the least minimum *P*-value, least number of true positives and the least average *P*-value followed by COPA, followed by *t*, followed by ORT, followed by TORT while *Z* has the highest maximum *P*-value, highest true positive rate and the highest average *P*-Value. Since OS has the least average *P*-value, it implies that OS performs better than the other methods. Based on this, OS has a higher detection power than the rest of the other methods followed by COPA, *t*, ORT, TORT and Z.

**Table 10** shows the ranking of the *P*-values computed by the various outlier techniques in an ascending order. The *P*-values and the Ranks were used in computing the False Discovery Rate FDR for the various outlier techniques

From **Table 11**, the true positives are the false null hypothesis (Type II error). These are the probabilities of accepting the null hypothesis given that the null hypothesis is false and should be rejected. These true positives are *P*-values that are greater than the given significant level of $\alpha = 0.05$. The blank cells in the table are the False Positives.

From **Table 12**, the false positives are the true null hypothesis (Type I error). These are the probabilities of rejecting the null hypothesis given that the null hypothesis is true and should be accepted. These False Positives are *P*-values that are less than the given significant level $\alpha = 0.05$. The blank spaces in the table are the false null hypothesis.

**Table 13** shows the FDR computed by the various outlier techniques. The *P*-values and the rank of the *P*-values were used in computing the False Discovery Rate (FDR). We can observed that all the outlier methods have equal minimum FDR of 0 (zero) except *Z*-Statistic with minimum FDR of 160.506. Modified *Z* has the highest maximum FDR of 160.506 followed by TORT = 114.838, OS = 84.819, ORT = 83.962, *t*-Statistic = 81.806 while COPA has the least maximum FDR of 79.25.

The Mean of the FDRs of the different outlier techniques were computed giving the following results: $Z = 160.506$, $t = 42.6748813$, COPA = 13.8631125, OS = 11.4017125, ORT = 45.64245 and TORT = 48.4732125. From results obtained, we observed that OS has the least average FDR and minimum FDR. Since OS has the least error rate, it implies that OS performs better than the other methods. Based on this, OS has a highest detection power with a smaller FDR followed by COPA, *t*, ORT, TORT and Z.

From **Figure 1**, we can see the performance of the FDRs of the various outlier methods. From the plot, we can see that the FDR of *Z* has the highest point at 160 constantly at the peak of the plot folllowed by the FDR of TORT, *t* and ORT. COPA has its highest point at the middle of the plot while OS has its points clustered at the floor of the plot. Based on these observations, we can see that OS performs better than the other methods in terms of having a smaller error rate (FDR) and therefore has the highest detection power follwed by COPA, ORT, *t*, TORT and Z.

## 3.2. Comparison Based on ROC Curves

The sensitivities were plotted against the specificities at different thresholds to compare the behaviour of the outlier methods. The ROC Curves were plotted for $n = 27$ and $k = 6$, $n = 27$ and $k = 10$, $n = 27$ and $k = 16$, $n = 27$ and $k = 25$. Where k is the numbers simulations. Larger area under the ROC curves indicates better sensitivity and specificity. An ROC curve along the diagonal line indicates a random-guess. The test result variable(s): *Z*, *t*, COPA, OS, ORT, TORT has at least one tie between the positive actual state group and the negative actual state group.

a) The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

**Table 10.** Summary of ranks of the *P*-values for various outlier techniques.

| *Z* | *T* | COPA | OS | ORT | TORT |
|---|---|---|---|---|---|
| 39.5 | 23.5 | 45 | 66 | 15 | 15.5 |
| 39.5 | 11 | 4 | 21.5 | 1.5 | 15.5 |
| 39.5 | 70.5 | 15.5 | 70 | 53 | 35 |
| 39.5 | 1.5 | 40.5 | 21.5 | 66 | 37 |
| 39.5 | 16 | 4 | 21.5 | 30 | 68 |
| 39.5 | 68 | 4 | 69 | 71 | 53 |
| 39.5 | 47.5 | 28 | 21.5 | 25 | 15.5 |
| 39.5 | 20.5 | 20 | 21.5 | 3 | 15.5 |
| 39.5 | 22 | 24 | 62 | 32 | 15.5 |
| 39.5 | 23.5 | 4 | 21.5 | 20 | 15.5 |
| 39.5 | 55 | 19 | 56 | 50 | 15.5 |
| 39.5 | 33.5 | 17 | 21.5 | 34 | 68 |
| 39.5 | 77.5 | 11.5 | 50 | 46 | 48 |
| 39.5 | 70.5 | 23 | 74 | 68 | 52 |
| 39.5 | 38.5 | 33 | 67 | 73 | 54 |
| 39.5 | 36.5 | 9 | 48.5 | 47 | 39 |
| 39.5 | 64 | 37 | 46 | 63 | 43 |
| 39.5 | 38.5 | 15.5 | 21.5 | 12 | 15.5 |
| 39.5 | 43.5 | 22 | 21.5 | 43 | 41 |
| 39.5 | 69 | 25 | 21.5 | 33 | 36 |
| 39.5 | 36.5 | 47 | 48.5 | 44 | 68 |
| 39.5 | 33.5 | 4 | 21.5 | 14 | 15.5 |
| 39.5 | 33.5 | 40.5 | 65 | 70 | 44 |
| 39.5 | 77.5 | 18 | 71 | 67 | 47 |
| 39.5 | 52.5 | 50 | 54 | 56 | 68 |
| 39.5 | 33.5 | 57 | 21.5 | 40 | 15.5 |
| 39.5 | 41 | 67 | 21.5 | 48 | 15.5 |
| 39.5 | 27.5 | 48 | 21.5 | 27 | 15.5 |
| 39.5 | 8.5 | 69 | 21.5 | 5 | 15.5 |
| 39.5 | 18 | 59 | 21.5 | 1.5 | 15.5 |
| 39.5 | 47.5 | 69 | 21.5 | 38 | 68 |
| 39.5 | 41 | 80 | 21.5 | 22 | 31 |
| 39.5 | 56 | 60 | 52 | 49 | 68 |
| 39.5 | 14.5 | 39 | 21.5 | 13 | 15.5 |
| 39.5 | 19 | 27 | 21.5 | 7 | 15.5 |
| 39.5 | 31 | 61 | 77.5 | 77.5 | 68 |
| 39.5 | 27.5 | 66 | 77.5 | 77.5 | 68 |
| 39.5 | 52.5 | 42 | 77.5 | 77.5 | 68 |
| 39.5 | 6 | 10 | 21.5 | 6 | 34 |

**Continued**

| | | | | | |
|---|---|---|---|---|---|
| 39.5 | 20.5 | 72 | 55 | 45 | 68 |
| 39.5 | 72 | 14 | 64 | 8 | 15.5 |
| 39.5 | 54 | 11.5 | 44 | 26 | 68 |
| 39.5 | 58 | 4 | 21.5 | 37 | 68 |
| 39.5 | 12.5 | 74 | 21.5 | 51 | 15.5 |
| 39.5 | 29 | 52 | 21.5 | 23 | 15.5 |
| 39.5 | 43.5 | 53 | 72 | 54 | 50 |
| 39.5 | 1.5 | 32 | 77.5 | 77.5 | 68 |
| 39.5 | 8.5 | 29 | 77.5 | 77.5 | 68 |
| 39.5 | 62.5 | 77 | 21.5 | 31 | 15.5 |
| 39.5 | 62.5 | 30 | 58 | 55 | 68 |
| 39.5 | 14.5 | 38 | 21.5 | 4 | 68 |
| 39.5 | 12.5 | 51 | 60 | 60 | 68 |
| 39.5 | 41 | 76 | 21.5 | 36 | 15.5 |
| 39.5 | 30 | 31 | 21.5 | 10 | 38 |
| 39.5 | 77.5 | 26 | 21.5 | 19 | 15.5 |
| 39.5 | 66 | 36 | 21.5 | 18 | 15.5 |
| 39.5 | 58 | 64 | 57 | 41 | 15.5 |
| 39.5 | 45.5 | 13 | 21.5 | 11 | 15.5 |
| 39.5 | 26 | 43 | 21.5 | 16 | 15.5 |
| 39.5 | 45.5 | 63 | 21.5 | 21 | 33 |
| 39.5 | 3.5 | 79 | 73 | 39 | 40 |
| 39.5 | 10 | 69 | 21.5 | 64 | 45 |
| 39.5 | 60 | 62 | 21.5 | 29 | 15.5 |
| 39.5 | 77.5 | 46 | 21.5 | 24 | 68 |
| 39.5 | 73 | 58 | 21.5 | 17 | 15.5 |
| 39.5 | 17 | 49 | 21.5 | 72 | 51 |
| 39.5 | 3.5 | 8 | 47 | 69 | 42 |
| 39.5 | 7 | 73 | 59 | 61 | 68 |
| 39.5 | 50 | 56 | 21.5 | 28 | 15.5 |
| 39.5 | 5 | 44 | 63 | 74 | 55 |
| 39.5 | 77.5 | 54.5 | 44 | 58.5 | 68 |
| 39.5 | 77.5 | 54.5 | 44 | 58.5 | 68 |
| 39.5 | 61 | 75 | 21.5 | 42 | 15.5 |
| 39.5 | 51 | 4 | 21.5 | 35 | 49 |
| 39.5 | 66 | 21 | 51 | 57 | 68 |
| 39.5 | 25 | 65 | 68 | 52 | 46 |
| 39.5 | 74 | 34 | 21.5 | 9 | 32 |
| 39.5 | 58 | 78 | 77.5 | 77.5 | 68 |
| 39.5 | 49 | 71 | 53 | 65 | 68 |
| 39.5 | 66 | 35 | 61 | 62 | 68 |

**Table 11.** Summary of calculated True Positives (TP) for various outlier techniques.

| TP of Z | TP of T | TP of COPA | TP of OS | TP of ORT | TP of TORT |
|---|---|---|---|---|---|
| 1 | 0.07 | 0.08 | 0.06 | 0.55 | 1 |
| 1 | 0.67 | 0.08 | 0.44 | 0.78 | 0.96 |
| 1 | 0.67 | 0.09 | 0.16 | 0.17 | 1 |
| 1 | 0.38 | 0.17 | 0.4 | 0.88 | 0.83 |
| 1 | 0.05 | 0.29 | 0.38 | 0.11 | 0.92 |
| 1 | 0.05 | 0.08 | 0.99 | 0.18 | 0.97 |
| 1 | 0.07 | 0.29 | 0.21 | 0.06 | 0.11 |
| 1 | 0.61 | 0.21 | 1 | 0.48 | 0.54 |
| 1 | 0.27 | 0.29 | 1 | 0.22 | 0.39 |
| 1 | 1 | 1 | 1 | 0.44 | 1 |
| 1 | 0.67 | 0.21 | 0.64 | 0.8 | 0.58 |
| 1 | 0.3 | 0.21 | 1 | 0.91 | 0.83 |
| 1 | 0.28 | 0.29 | 1 | 0.45 | 1 |
| 1 | 0.66 | 0.05 | 0.93 | 0.7 | 1 |
| 1 | 0.3 | 0.36 | 0.28 | 0.35 | 1 |
| 1 | 0.34 | 0.41 | 1 | 0.21 | 1 |
| 1 | 0.67 | 0.12 |  | 0.36 | 1 |
| 1 | 0.28 | 0.12 |  | 0.86 | 1 |
| 1 | 0.27 | 0.65 |  | 0.79 | 1 |
| 1 | 0.27 | 0.1 |  | 0.62 | 1 |
| 1 | 1 | 0.61 |  | 0.3 | 1 |
| 1 | 0.55 | 0.27 |  | 0.46 | 0.84 |
| 1 | 0.27 | 0.06 |  | 0.12 | 1 |
| 1 | 0.33 | 0.25 |  | 0.28 | 1 |
| 1 | 0.1 | 0.91 |  | 0.08 | 1 |
| 1 | 0.38 | 0.29 |  | 0.47 | 1 |
| 1 | 0.33 | 0.24 |  | 1 | 1 |
| 1 | 0.62 | 0.08 |  | 1 | 0.12 |
| 1 | 1 | 0.18 |  | 1 | 0.61 |
| 1 | 1 | 0.09 |  | 0.42 | 1 |
| 1 | 0.65 | 0.4 |  | 0.11 | 0.89 |
| 1 | 0.43 | 0.16 |  | 0.26 | 0.52 |
| 1 | 0.67 | 0.06 |  | 0.48 | 1 |
| 1 | 0.07 | 0.15 |  | 0.08 | 0.97 |
| 1 | 0.72 | 0.15 |  | 0.57 | 1 |
| 1 | 0.63 | 0.6 |  | 1 | 1 |
| 1 | 0.39 | 0.28 |  | 1 | 0.83 |
| 1 | 0.67 | 0.88 |  | 0.17 | 1 |
| 1 | 0.23 | 0.31 |  | 0.58 | 0.82 |

**Continued**

| | | | |
|---|---|---|---|
| 1 | 0.1 | 0.66 | 1 |
| 1 | 0.55 | 0.26 | 1 |
| 1 | 0.05 | 0.06 | 1 |
| 1 | 0.68 | 0.31 | |
| 1 | 0.58 | 0.07 | |
| 1 | 0.63 | 0.29 | |
| 1 | 0.13 | 0.73 | |
| 1 | 0.34 | 0.15 | |
| 1 | 0.65 | 0.08 | |
| 1 | 0.65 | 0.91 | |
| 1 | 0.33 | 0.83 | |
| 1 | 0.15 | 0.68 | |
| 1 | 1 | 0.13 | |
| 1 | 0.67 | 0.98 | |
| 1 | 0.63 | 0.64 | |
| 1 | 0.37 | 0.64 | |
| 1 | 0.07 | 0.32 | |
| 1 | 0.37 | 0.23 | |
| 1 | 0.63 | 0.64 | |
| 1 | 1 | 0.54 | |
| 1 | 0.7 | 1 | |
| 1 | 0.4 | 0.75 | |
| 1 | | 0.7 | |
| 1 | | | |
| 1 | | | |
| 1 | | | |
| 1 | | | |
| 1 | | | |
| 1 | | | |
| 1 | | | |
| 1 | | | |
| 1 | | | |
| 1 | | | |
| 1 | | | |
| 1 | | | |
| 1 | | | |
| 1 | | | |
| 1 | | | |
| 1 | | | |
| 1 | | | |
| 1 | | | |

**Table 12.** Summary of calculated False Positives (FP) for various outlier techniques.

| FP of *Z* | FP of *T* | FP of COPA | FP of OS | FP of ORT | FP of TORT |
|---|---|---|---|---|---|
| 0 | 0.02 | 0 | 0 | 0.03 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0.03 | 0.04 | 0 | 0 | 0 |
| 0 | 0.01 | 0 | 0 | 0.02 | 0.02 |
| 0 | 0.04 | 0 | 0 | 0.03 | 0 |
| 0 | 0.02 | 0.01 | 0.04 | 0 | 0 |
| 0 | 0.04 | 0 | 0 | 0 | 0 |
| 0 | 0.01 | 0.01 | 0.01 | 0.02 | 0 |
| 0 | 0.02 | 0 | 0 | 0.01 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0.01 | 0 | 0 | 0.01 | 0.02 |
| 0 | 0.02 | 0 | 0 | 0.02 | 0 |
| 0 | 0.02 | 0.01 | 0 | 0.04 | 0 |
| 0 | 0 | 0.01 | 0 | 0.02 | 0 |
| 0 | 0.01 | 0 | 0 | 0.03 | 0 |
| 0 | 0.03 | 0.03 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0.03 | 0 |
| 0 | 0.01 | 0 | 0 | 0.01 | 0 |
| 0 | 0 | 0.01 | 0 | | 0 |
| 0 | | 0 | 0 | | 0 |
| 0 | | 0.04 | 0 | | 0 |
| 0 | | 0 | 0 | | 0 |
| 0 | | 0.04 | 0 | | 0 |
| 0 | | 0.01 | 0 | | 0 |
| 0 | | 0 | 0 | | 0 |
| 0 | | 0 | 0 | | 0 |
| 0 | | 0 | 0 | | 0.02 |
| 0 | | 0 | 0 | | 0 |
| 0 | | 0.01 | 0 | | 0 |
| 0 | | 0.01 | 0 | | 0 |
| 0 | | 0.01 | 0.04 | | 0 |
| 0 | | 0.04 | 0 | | 0 |
| 0 | | 0.01 | 0 | | 0 |
| 0 | | 0.01 | 0 | | 0 |
| 0 | | 0.03 | 0 | | 0 |
| 0 | | 0 | 0 | | 0 |
| 0 | | 0 | 0.01 | | 0 |
| 0 | | 0 | 0 | | 0 |
| 0 | | 0 | 0.02 | | |

**Continued**

| | | |
|---|---|---|
| 0 | 0.02 | 0 |
| 0 | 0.03 | 0 |
| 0 | | 0 |
| 0 | | 0 |
| 0 | | 0.01 |
| 0 | | 0 |
| 0 | | 0 |
| 0 | | 0 |
| 0 | | 0 |
| 0 | | 0 |
| 0 | | 0 |
| 0 | | 0 |
| 0 | | 0 |
| 0 | | 0 |
| 0 | | 0.01 |
| 0 | | 0 |
| 0 | | 0.04 |
| 0 | | 0 |
| 0 | | 0 |
| 0 | | 0 |
| 0 | | 0 |
| 0 | | 0 |
| 0 | | 0 |
| | | 0 |
| 0 | | 0.03 |
| 0 | | |
| 0 | | |
| 0 | | |
| 0 | | |
| 0 | | |
| 0 | | |
| 0 | | |
| 0 | | |
| 0 | | |
| 0 | | |
| 0 | | |
| 0 | | |
| 0 | | |
| 0 | | |

**Table 13.** Summary of computed False Discovery Rate (FDR) for various outlier techniques.

| FDR OF *Z* | FDR OF *T* | FDR OF COPA | FDR OF OS | FDR OF ORT | FDR OF TORT |
|---|---|---|---|---|---|
| 160.506 | 18.885 | 11.271 | 15.37 | 12.68 | 0 |
| 160.506 | 11.527 | 0 | 0 | 0 | 0 |
| 160.506 | 60.252 | 0 | 36.229 | 65.792 | 3.427 |
| 160.506 | 0 | 6.262 | 0 | 74.927 | 93.235 |
| 160.506 | 11.888 | 0 | 0 | 35.927 | 114.838 |
| 160.506 | 62.468 | 0 | 34.916 | 78.58 | 0 |
| 160.506 | 50.72 | 2.264 | 0 | 27.896 | 0 |
| 160.506 | 15.463 | 0 | 0 | 0 | 0 |
| 160.506 | 14.409 | 2.642 | 4.09 | 35.663 | 0 |
| 160.506 | 18.885 | 0 | 0 | 19.02 | 0 |
| 160.506 | 70.316 | 0 | 1.132 | 60.864 | 93.235 |
| 160.506 | 51.099 | 0 | 0 | 41.024 | 109.629 |
| 160.506 | 81.806 | 0 | 0 | 60.643 | 112.169 |
| 160.506 | 60.252 | 2.757 | 84.819 | 74.588 | 113.885 |
| 160.506 | 49.403 | 1.921 | 19.872 | 79.033 | 17.882 |
| 160.506 | 48.636 | 0 | 0 | 60.702 | 79.619 |
| 160.506 | 65.381 | 5.141 | 0 | 70.444 | 0 |
| 160.506 | 49.403 | 0 | 0 | 10.567 | 60.307 |
| 160.506 | 49.554 | 0 | 0 | 51.605 | 3.522 |
| 160.506 | 61.562 | 2.536 | 0 | 40.345 | 93.235 |
| 160.506 | 48.636 | 10.791 | 0 | 51.873 | 0 |
| 160.506 | 51.099 | 0 | 0 | 13.586 | 83.573 |
| 160.506 | 51.099 | 6.262 | 5.852 | 77.891 | 111.962 |
| 160.506 | 81.806 | 0 | 39.29 | 74.755 | 93.235 |
| 160.506 | 66.419 | 11.412 | 0 | 70.193 | 0 |
| 160.506 | 51.099 | 18.909 | 0 | 47.55 | 0 |
| 160.506 | 51.029 | 27.442 | 0 | 60.758 | 0 |
| 160.506 | 23.055 | 10.567 | 0 | 28.178 | 0 |
| 160.506 | 7.459 | 26.646 | 0 | 0 | 0 |
| 160.506 | 14.089 | 22.566 | 0 | 0 | 93.235 |
| 160.506 | 50.72 | 26.646 | 0 | 46.716 | 0 |
| 160.506 | 51.029 | 79.25 | 0 | 23.055 | 93.235 |
| 160.506 | 70.193 | 22.19 | 0 | 60.812 | 0 |
| 160.506 | 8.745 | 6.503 | 0 | 9.754 | 0 |
| 160.506 | 13.347 | 2.348 | 0 | 9.057 | 93.235 |
| 160.506 | 47.039 | 21.826 | 81.806 | 81.806 | 93.235 |
| 160.506 | 23.055 | 27.858 | 81.806 | 81.806 | 93.235 |
| 160.506 | 66.419 | 7.548 | 81.806 | 81.806 | 0 |
| 160.506 | 10.567 | 0 | 0 | 0 | 93.235 |
| 160.506 | 15.463 | 31.7 | 0 | 59.173 | 0 |

**Continued**

| | | | | | |
|---|---|---|---|---|---|
| 160.506 | 59.878 | 0 | 3.963 | 7.925 | 93.235 |
| 160.506 | 68.096 | 0 | 0 | 26.823 | 93.235 |
| 160.506 | 68.866 | 0 | 0 | 44.551 | 0 |
| 160.506 | 10.144 | 35.127 | 0 | 59.671 | 0 |
| 160.506 | 28.421 | 14.631 | 0 | 22.052 | 106.512 |
| 160.506 | 49.554 | 14.355 | 56.356 | 66.922 | 93.235 |
| 160.506 | 0 | 1.981 | 81.806 | 81.806 | 93.235 |
| 160.506 | 7.459 | 2.186 | 81.806 | 81.806 | 0 |
| 160.506 | 65.936 | 53.519 | 0 | 34.768 | 93.235 |
| 160.506 | 65.936 | 2.113 | 1.093 | 66.858 | 93.235 |
| 160.506 | 8.745 | 6.674 | 0 | 0 | 93.235 |
| 160.506 | 10.144 | 12.431 | 2.113 | 69.74 | 0 |
| 160.506 | 51.029 | 50.887 | 0 | 45.789 | 3.337 |
| 160.506 | 31.7 | 2.045 | 0 | 12.68 | 0 |
| 160.506 | 81.806 | 2.438 | 0 | 20.021 | 0 |
| 160.506 | 64.361 | 5.283 | 0 | 14.089 | 0 |
| 160.506 | 68.866 | 26.747 | 1.112 | 47.937 | 0 |
| 160.506 | 51.556 | 0 | 0 | 11.527 | 0 |
| 160.506 | 17.069 | 8.847 | 0 | 11.888 | 0 |
| 160.506 | 51.556 | 25.159 | 0 | 21.133 | 19.02 |
| 160.506 | 0 | 73.03 | 80.77 | 47.144 | 85.942 |
| 160.506 | 6.34 | 26.646 | 0 | 72.316 | 0 |
| 160.506 | 66.57 | 24.542 | 0 | 32.793 | 93.235 |
| 160.506 | 81.806 | 11.026 | 0 | 21.133 | 0 |
| 160.506 | 60.795 | 19.676 | 0 | 11.188 | 110.639 |
| 160.506 | 11.188 | 11.645 | 0 | 80.131 | 78.495 |
| 160.506 | 0 | 0 | 0 | 76.264 | 93.235 |
| 160.506 | 9.057 | 34.74 | 1.075 | 70.675 | 0 |
| 160.506 | 50.72 | 18.114 | 0 | 29.436 | 111.815 |
| 160.506 | 0 | 8.645 | 4.025 | 83.962 | 93.235 |
| 160.506 | 81.806 | 17.45 | 0 | 69.361 | 93.235 |
| 160.506 | 81.806 | 17.45 | 0 | 69.361 | 0 |
| 160.506 | 67.557 | 50.72 | 0 | 48.305 | 107.392 |
| 160.506 | 53.455 | 0 | 0 | 41.663 | 93.235 |
| 160.506 | 64.361 | 0 | 0 | 71.186 | 113.017 |
| 160.506 | 17.752 | 27.311 | 26.106 | 65.838 | 0 |
| 160.506 | 61.686 | 3.729 | 0 | 7.044 | 93.235 |
| 160.506 | 68.866 | 71.528 | 81.806 | 81.806 | 93.235 |
| 160.506 | 50.461 | 27.682 | 0 | 73.154 | 93.235 |
| 160.506 | 64.361 | 5.434 | 3.118 | 71.581 | 0 |
| 160.506 | 42.674813 | 13.863113 | 11.401713 | 45.64245 | 48.473213 |

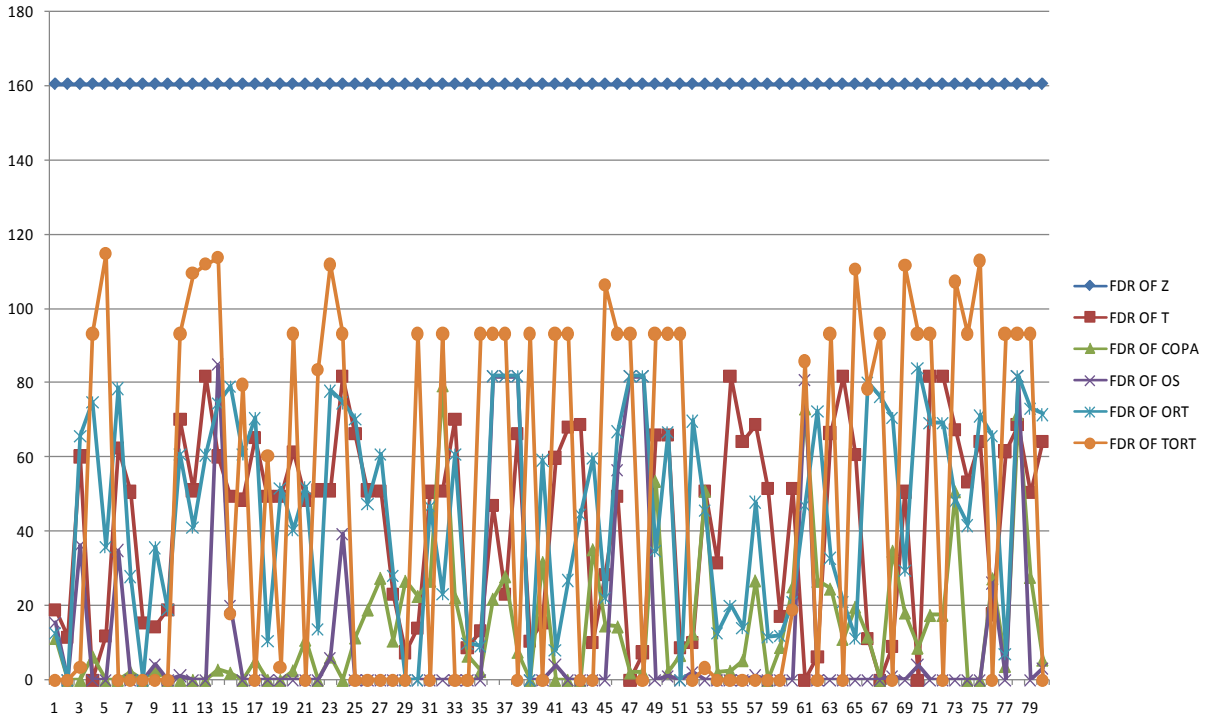


**Figure 1.** Plot of the FDRs.

b) Null hypothesis: Significant/True Area under the ROC Curve (AUC) = 0.5.

From the **Figure 2**, we can observe that for a smaller $k = 6$, $T$ and OS have a larger Area Under the ROC Curve AUC with values 0.813 and 0.750 which indicate better sensitivity and specificity and are significant. COPA and ORT have equal AUC with point 0.563 followed by $Z$ with point 0.500 while TORT has the least AUC of 0.375 which is not significant.

**Table 14** shows the ROC curves analysis for $n = 27$ and $k = 6$, the area under the ROC curve (AUC) and the confidence interval for all the test variables.

From **Figure 3**, we can observe that for a smaller $k = 10$, $T$ and OS have a larger Area Under the ROC Curve AUC with values 0.781 and 0.719 which indicate better sensitivity and specificity and are significant. COPA and ORT have equal significant AUC with point 0.594 followed by $Z$ with point 0.500 which is on the refference line while TORT has the least AUC of 0.438 which is not significant.

**Table 15** shows the ROC curves analysis for $n = 27$ and $k = 10$, the area under the ROC curve (AUC) and the confidence interval for all the test variables.

From **Figure 4**, we can observe that for bigger $k = 16$, $T$ and OS have a larger Area Under the ROC Curve AUC with values 0.804 and 0.732 which indicate better sensitivity and specificity and are more significant. COPA and ORT have better significant AUC with points 0.536 and .0518 followed by $Z$ with point 0.500 which is on the refference line while TORT has the least AUC of 0.429 which is not significant.

**Table 16** shows the ROC curves analysis for $n = 27$ and $k = 16$, the area under the ROC curve (AUC) and the confidence interval for all the test variables.

From **Figure 5**, we can observe that for a bigger $k = 25$, $t$ and OS have a larger Area Under the ROC Curve AUC with values 0.828 and 0.734 which indicate better sensitivity and specificity and are more significant. COPA has a better significant AUC with points 0.563 followed by $Z$ and ORT which have equal AUC with point 0.500 while TORT has the least AUC of 0.438 which is not significant.

**Table 17** shows the ROC curves analysis for $n = 27$ and $k = 20$, the area under the ROC curve (AUC) and the confidence interval for all the test variables.

**Table 18** is a summary of all the findings in the analysis for all the outlier techniques on the bases of their $P$-values, false positives, true positives, false discovery rates and their corresponding ROC curves.

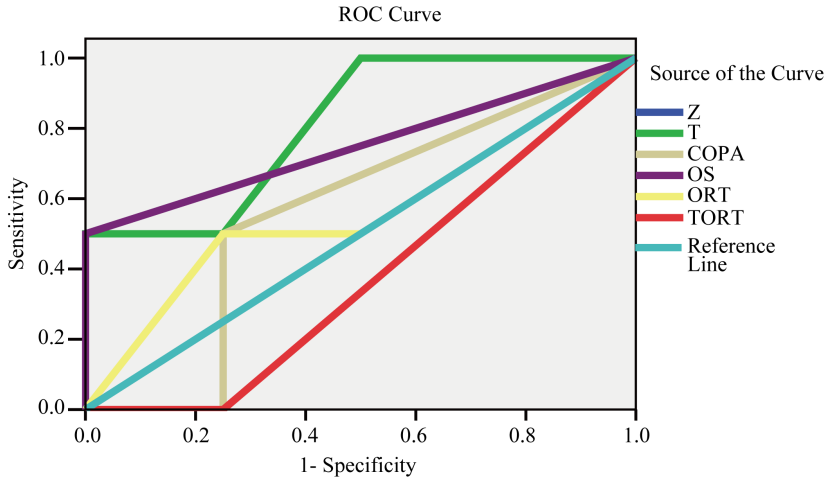


**Figure 2.** ROC curve when $n = 27$ and $k = 6$.

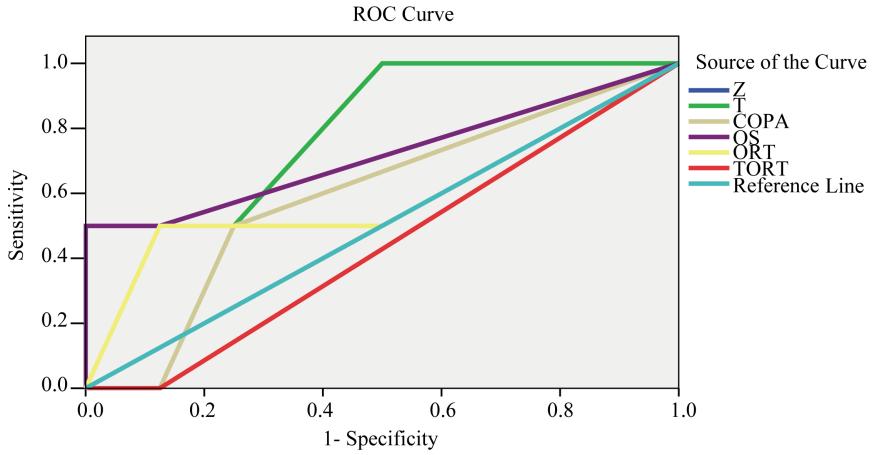


**Figure 3.** ROC curve when $n = 27$ and $k = 10$.

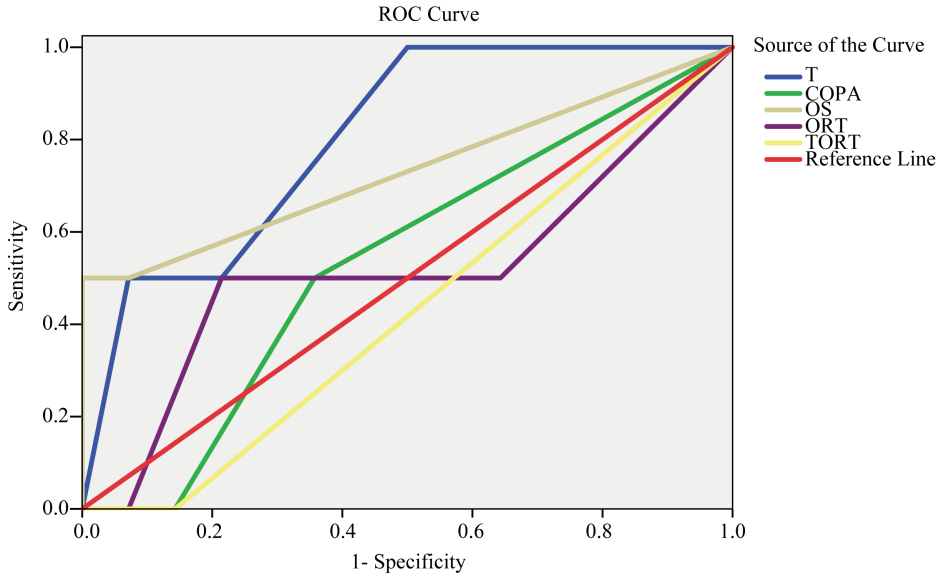


**Figure 4.** ROC curve when $n = 27$ and $k = 16$.

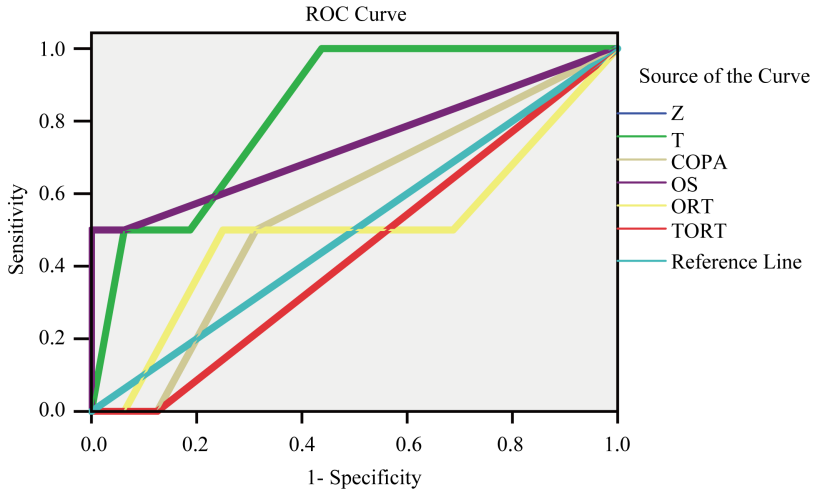


**Figure 5.** ROC curve when $n = 27$ and $k = 25$.

**Table 14.** ROC curve when $n = 27$ and $k = 6$ area under the curve.

| Test Result Variable(s) | Area | Std. Error[a] | Asymptotic Sig.[b] | Asymptotic 95% Confidence Interval | |
|---|---|---|---|---|---|
| | | | | Lower Bound | Upper Bound |
| *Z* | 0.500 | 0.270 | 1.000 | 0.000 | 1.000 |
| *T* | 0.813 | 0.198 | 0.247 | 0.000 | 1.000 |
| COPA | 0.563 | 0.261 | 0.817 | 0.000 | 1.000 |
| OS | 0.750 | 0.255 | 0.355 | 0.000 | 1.000 |
| ORT | 0.563 | 0.282 | 0.817 | 0.000 | 1.000 |
| TORT | 0.375 | 0.246 | 0.643 | 0.000 | 1.000 |

**Table 15.** ROC curve when $n = 27$ and $k = 10$ area under the curve.

| Test Result Variable(s) | Area | Std. Error[a] | Asymptotic Sig.[b] | Asymptotic 95% Confidence Interval | |
|---|---|---|---|---|---|
| | | | | Lower Bound | Upper Bound |
| *Z* | 0.500 | 0.239 | 1.000 | 0.031 | 0.969 |
| *T* | 0.781 | 0.166 | 0.240 | 0.000 | 10.000 |
| COPA | 0.594 | 0.229 | 0.695 | 0.076 | 10.000 |
| OS | 0.719 | 0.251 | 0.361 | 0.000 | 10.000 |
| ORT | 0.594 | 0.273 | 0.695 | 0.000 | 10.000 |
| TORT | 0.438 | 0.223 | 0.794 | 0.001 | 0.874 |

**Table 16.** ROC curve when $n = 27$ and $k = 16$ area under the curve.

| Test Result Variable(s) | Area | Std. Error[a] | Asymptotic Sig.[b] | Asymptotic 95% Confidence Interval | |
|---|---|---|---|---|---|
| | | | | Lower Bound | Upper Bound |
| *Z* | 0.500 | 0.225 | 1.000 | 0.059 | 0.941 |
| *T* | 0.804 | 0.147 | 0.177 | 0.000 | 1.000 |
| COPA | 0.536 | 0.207 | 0.874 | 0.130 | 0.942 |
| OS | 0.732 | 0.240 | 0.302 | 0.000 | 1.000 |
| ORT | 0.518 | 0.259 | 0.937 | 0.000 | 1.000 |
| TORT | 0.429 | 0.202 | 0.751 | 0.033 | 0.825 |

**Table 17.** ROC curve when $n = 27$ and $k = 25$ area under the curve.

| Test Result Variable(s) | Area | Std. Error[a] | Asymptotic Sig.[b] | Asymptotic 95% Confidence Interval | |
|---|---|---|---|---|---|
| | | | | Lower Bound | Upper Bound |
| *Z* | 0.500 | 0.222 | 1.000 | 0.064 | 0.936 |
| *T* | 0.828 | 0.130 | 0.140 | 0.000 | 10.000 |
| COPA | 0.563 | 0.208 | 0.779 | 0.155 | 0.970 |
| OS | 0.734 | 0.239 | 0.292 | 0.000 | 1.000 |
| ORT | 0.500 | 0.260 | 10.000 | 0.000 | 1.000 |
| TORT | 0.438 | 0.202 | 0.779 | 0.042 | 0.833 |

**Table 18.** Summary of findings for the various outlier techniques.

| DESCRIPTIVES / OUTLIER TECHNIQUES | Modified *Z* | *t*-Statistic | COPA | OS | ORT | TORT |
|---|---|---|---|---|---|---|
| Maximum *P*-Value | 1 | 1 | 1 | 1 | 1 | 1 |
| Minimum *P*-Value | 1 | 0 | 0 | 0.0001 | 0.0001 | 0.0001 |
| Mean *P*-Value | 1 | 0.360 | 0.145 | 0.134 | 0.389 | 0.473 |
| True Positives | 80 | 61 | 39 | 16 | 62 | 42 |
| False Positives | 0 | 19 | 41 | 64 | 18 | 38 |
| Minimum FDR | 160.506 | 0 | 0 | 0 | 0 | 0 |
| Maximum FDR | 160.506 | 81.806 | 79.25 | 84.819 | 83.962 | 114.838 |
| Mean FDR | 160.506 | 42.675 | 13.863 | 11.402 | 45.642 | 48.473 |
| True(Significant) AUC = 0.5 AUC of ROC curve for $k = 6$ | 0.500 | 0.813 | 0.563 | 0.750 | 0.563 | 0.375 |
| AUC of ROC curve for $k = 10$ | 0.500 | 0.781 | 0.594 | 0.719 | 0.594 | 0.438 |
| AUC of ROC curve for $k = 16$ | 0.500 | 0.804 | 0.536 | 0.732 | 0.518 | 0.429 |
| AUC of ROC curve for $k = 25$ | 0.500 | 0.828 | 0.563 | 0.734 | 0.500 | 0.438 |

## 4. Conclusion

The performance of the various outlier methods—*Z*, *T*, COPA, OS, ORT and TORT has been statistically studied using simulated data to evaluate which of these methods has the highest power of detecting and handling outliers in terms of their *P*-Values, true positives, false positives, false discovery rate FDR and their corresponding ROC curves.

The result of their *P*-values showed that all the outlier methods have equal maximum *P*-value. Modified *Z* has the highest minimum *P*-Value followed by *T* and COPA while OS, ORT and TORT have the least minimum *P*-Value. Modified *Z* has the highest true positives rate followed by ORT, *t*-Statistic, TORT, COPA, while OS has the least true positives rate. *Z* has the highest average *P*-Value followed by TORT, ORT, T, COPA while OS has the least average *P*-Value. Based on these results, OS performed better than the methods followed by COPA, *T*, ORT, TORT and *Z* in terms of their *P*-Values. When comparison was made on the FDRs, OS also performs the best by having the smallest FDR followed by COPA, *T*, ORT, TORT and *Z*.

In terms of their ROC curves, for a smaller $k = 6$ and 10, *T* and OS have the largest Area Under the ROC Curve AUC which indicate a better sensitivity and specificity and are significant. COPA and ORT have equal significant AUC followed by *Z* with insignificant AUC while TORT has the least AUC which is not significant. Also for larger $k = 16$ and 25, *T* and OS still have the largest Area Under the ROC Curve AUC which indicate better sensitivity and specificity and are more significant. COPA and ORT have better significant AUC followed by *Z* with insignificant AUC while TORT has the least AUC which is still not significant. Based on the above results so far obtained from this analysis, it is obvious that the Outlier Sum Statistic OS has more power of detecting and handling outliers with a smaller False Discovery Rate (FDR) followed by COPA, *T*, ORT, TORT and *Z*.

## References


[1] Grubbs, F.E. (1969) Procedures for Detecting Outlying Observations in Samples. *Technometrics*, **11**, 1-21. http://dx.doi.org/10.1080/00401706.1969.10490657

[2] Hawkins, D. (1980) Identification of Outliers. Chapman and Hall, Kluwer Academic Publishers, Boston/Dordrecht/ London.

[3] Aggarwal, C.C. (2005) On Abnormality Detection in Spuriously Populated Data Streams. SIAM Conference on Data Mining. Kluwer Academic Publishers Boston/Dordrech/London.

[4] Barnett, V. and Lewis, T. (1994) Outliers in Statistical Data. 3rd Edition, John Wiley & Sons, Kluwer Academic Publishers, Boston/Dordrecht/London.

[5] Dudoit, S., Yang, Y., Callow, M. and Speed, T. (2002) Statistical Methods for Identifying Differentially Expressed Genes in Replicated DNA Microarray Experiments. *Statistica Sinica*, **12**, 111-139.

[6] Troyanskaya, O.G., Garber, M.E., Brown, P.O., Botstein, D. and Altman, R.B. (2002) Nonparametric Methods for Identifying Differentially Expressed Genes in Microarray Data. *Bioinformatics*, **18**, 1454-1461. http://dx.doi.org/10.1093/bioinformatics/18.11.1454

[7] Tomlins, S., Rhodes, D., Perner, S., Dhanasekaran, S., Mehra, R., Sun, X., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., *et al*. (2005) Recurrent Fusion of *TMPRSS*2 and ETS Transcription Factor Genes in Prostate Cancer. *Science*, **310**, 644-648. http://dx.doi.org/10.1126/science.1117679

[8] Efron, B., Tibshirani, R., Storey, J. and Tusher, V. (2001) Empirical Bayes Analysis of a Microarray Experiment. *Journal of the American Statistical Association*, **96**, 1151-1160. http://dx.doi.org/10.1198/016214501753382129

[9] Iglewicz, B. and Hoaglin, D.C. (2010) Detection of Outliers. Engineering Statistical Handbook, 1.3.5.17. Database Systems Group.

[10] Lyons-Weiler, J., Patel, S., Becich, M. and Godfrey, T. (2004) Tests for Finding Complex Patterns of Differential Expression in Cancers: Towards Individualized Medicine. *Bioinformatics*, **5**, 1-9.

[11] Tibshirani, R. and Hastie, R. (2006) Outlier Sums Statistic for Differential Gene Expression Analysis. *Biostatistics*, **8**, 2-8. http://dx.doi.org/10.1093/biostatistics/kxl005

[12] Benjamini, Y. and Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B*, **57**, 289-300.

[13] Wu, B. (2007) Cancer Outlier Differential Gene Expression Detection. *Biostatistics*, **8**, 566-575. http://dx.doi.org/10.1093/biostatistics/kxl029

[14] Luo, J. (2012) Truncated Outlier Robust T-Statistic for Outlier Detection. *Open Journal of Statistics*, **2**, 120-123. http://dx.doi.org/10.4236/ojs.2012.21013

[15] Fonseca, R., Barlogie, B., Bataille, R., Bastard, C., Bergsagel, P.L., Chesi, M., *et al*. (2004) Genetics and Cytogenetics of Multiple Myelom: A Workshop Report. *Cancer Research*, **64**, 1546- 1558. http://dx.doi.org/10.1158/0008-5472.CAN-03-2876

[16] MacDonald, J.W. and Ghosh, D. (2006) Copa—Cancer Outlier Profile Analysis. *Bioinformatics*, **22**, 2950-2951. http://dx.doi.org/10.1093/bioinformatics/btl433

[17] Hu, J. (2008) Cancer Outlier Detection Based on Likelihood Ratio Test. *Bioinformatics*, **24**, 2193-2199. http://dx.doi.org/10.1093/bioinformatics/btn372

[18] Lian, H. (2008) MOST: Detecting Cancer Differential Gene Expression. *Biostatistics*, **9**, 411-418. http://dx.doi.org/10.1093/biostatistics/kxm042

[19] Ghosh, D. (2009) Genomic Outlier Profile Analysis: Mixture Models, Null Hypotheses, and Nonparametric Estimation. *Biostatistics*, **10**, 60-69. http://dx.doi.org/10.1093/biostatistics/kxn015

[20] Chen, L.A., Chen, D.T. and Chan, W. (2010) The Distribution-Based *P*-value for the Outlier Sum in Differential Gene Expression Analysis. *Biometrika*, **97**, 246-253. http://dx.doi.org/10.1093/biomet/asp075

[21] Ghosh, D. (2010) Discrete Nonparametric Algorithms for Outlier Detection with Genomic Data. *Journal of Biopharmaceutical Statistics*, **20**, 193-208. http://dx.doi.org/10.1080/10543400903572704

[22] Filmoser, P., Maronna, R. and Werner, M. (2008) Outlier Identification in High Dimensions. *Computational Statistics and Data Analysis*, **52**, 1694-1711.

[23] Mori, K., Oura, T., Noma, H. and Matsui, S. (2013) Cancer Outlier Analysis Based on Mixture Modeling of Gene Expression Data. *Computational and Mathematical Methods in Medicine*, **2013**, Article ID: 693901. http://dx.doi.org/10.1155/2013/693901

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.