Zanco Journal of Human Sciences (ZJHS)
ISSN (Print):2218-0222
ISSN (Online): 2412-396X

Vol.28 Issue 12024
ID No.1334 (PP 140 - 157)
https://doi.org/10.21271/zjhs.28.2.11

Research Article

Construction Robust -Chart and compare it with Hotelling's T2-Chart

Bekhal Samad Sedeeq*
Zainab Abdulla Muhammad*
Israa Muayad Ali*
Taha Hussein Ali*

* Department of Statistics and Informatics, College of Administration and Economics, Salahaddin University-Erbil
bikhal.sedeeq@su.edu.krd
zainab.muhammad@su.edu.krd
israa.ali@su.edu.krd
taha.ali@su.edu.krd

Reiceved ‖ 02/07/2023
Accepted ‖ 10/08/2023
Published ‖ 15/02/2024

Keywords:
Multivariate Quality Control Charts,
Robust,
Outliers,
T2-Chart.

Abstract

This paper proposes a new multivariate chart corresponding to a T2- chart robust to outliers using three methods, namely the Rousseuw and Leroy algorithm, Maronna and Zamar, and the family of "concentration algorithms" by Olive and Hawkins. Then the comparison between the proposed and classical method of the researcher Shewhart depending on the total variance (trace variance matrix) and the general variance (determinant of the variance matrix) to obtain the most efficient paintings against outliers through simulation and real data and using a program in MATLAB language designed for this purpose. The study concluded that the proposed charts dealt with the problem of the influence of outliers and were more efficient than the classical method.

About the Journal
ZANCO Journal of Humanity Sciences (ZJHS) is an international, multi-disciplinary, peer-reviewed, double-blind and open-access journal that enhances research in all fields of basic and applied sciences through the publication of high-quality articles that describe significant and novel works; and advance knowledge in a diversity of scientific fields.
https://zancojournal.su.edu.krd/index.php/JAHS/about

1. Introduction

Quality Control Charts have historically been used to monitor product quality in a production or manufacturing environment. Their general purpose is to provide information that can be used to uncover discrepancies or systematic patterns by comparing expected variance and observed variance. In a production environment, it is important to improve product quality and productivity to maximize a company's profits (Deming, 1982)

Many different variations of control charts can be used to detect when processes go out of control. The most common and easily interpretable of these is the Shewart control chart (Ali et al 2017). These charts, named after Walter Shewart, were created from the assumption that every process that has variation can be understood and statistically monitored. A Shewart chart includes three horizontal lines, a center line, an upper limit, and a lower limit, and is the basis for all control charts (Ali and Esraa, 2016). The center line serves as a baseline and is typically the expected value or the mean value, while the upper and lower limits are depicted by baselines and are evenly spaced below and above the baseline. (Montgomery, 2009)

On another side, it is a fact of life that most data are naturally multivariate (or Bivariate). Hotelling in 1947 introduced a statistic that uniquely lends itself to plotting multivariate observations. This statistic, appropriately named Hotelling's T^2, is a scalar that combines information from the dispersion and mean of several variables. Because computations are laborious and fairly complex and require some knowledge of matrix algebra, acceptance of multivariate control charts by industry was slow and hesitant, In this research. Quality standards may be any one or a combination of attributes and variables of the product being manufactured. The attributes will include performance, reliability, appearance, commitment to delivery time, etc., variables may be some measurement variables like, length, width, height, diameter, surface finish, etc. Most of the above characteristics are related to products. Similarly, some of the quality characteristics of services are meeting promised due dates, safety, comfort, security, less waiting time, and so forth. So, the various dimensions of quality are performance, features, reliability, conformance, durability, serviceability, aesthetics, perceived quality, safety, comfort, security, commitment to due dates, less waiting time, etc. (Kovach, 2007)

Outliers are extreme values that stand out greatly from the overall pattern of values in a dataset or graph. Outliers are an important part of a dataset. They can hold useful information about your data. Outliers can give helpful insights into the data you're studying, and they can affect statistical results. This can potentially help you discover inconsistencies and detect any errors in your statistical processes (Ali et al 2019). So, knowing how to find outliers in a dataset will help you better understand your data. Outlier in Statistics is an extremely high or extremely low data point relative to the nearest data point and the rest of the neighboring co-existing values in a data graph or dataset you're working with. An outlier is an observation that lies an abnormal distance from other values in a random sample from a population. (Charu, 2017) (Kareem et al, 2020).

The presence of a small proportion of outliers in a sample can have a large distorting influence on the sample mean and the sample variance. It is well-known that these classical estimators, optimal under the normality assumption, are extremely sensitive to atypical observations in the data. In robust statistics methods are developed that are resistant to outliers in the data. Robust statistics seek to provide methods that emulate popular statistical methods but are not unduly affected by outliers or other small departures from model assumptions. In statistics, (Ali and Saleh 2022) classical estimation methods rely heavily on assumptions that are often not met in practice. In particular, it is often assumed that the data errors are normally distributed, at least approximately, or that the central limit theorem can be relied on to produce normally distributed estimates. Unfortunately, when there are outliers in the data, classical estimators often have very poor performance, when judged using the breakdown point and the influence function, (Farcomeni and Greco 2021).

2. Quality Control Chart

Quality Control has been a significant topic in industry since 1924 when Walter Shewhart published his first control chart. Managers immediately recognized the need to improve the quality of their products by improving the consistency of the manufacturing process. Until recently the primary focus has been on monitoring the quality of a process by observing a single variable/attribute over time and the relation of this quality characteristic to a set of predetermined criteria called control limits. Usually, the process is classified as in-control if the observed variable is within the control limits and out of control if it is outside of the control limits. A signal occurs when a quality characteristic's observed value at a point in time is beyond the predetermined control limits. When all quality characteristics in the process are deemed in control, the manager can feel secure that the monitored process is consistent with past performance. Note that a process being in control does not necessarily indicate that a high-quality product is being produced, but that the quality of the product is consistent with what has been produced historically. At a point when the control chart signals and the process is assumed out of control, the quality practitioner should investigate the process to determine what has changed in its operation. (Thomas, 2002)

2.1. Univariate Control Chart

The initial Shewhart chart, designed to monitor sample means, $\overline{X}$, of a process has developed through the implementation of runs rules and the availability of powerful computing facilities. It has also set the standard for a class of more elaborate control charts like the exponentially weighted moving average (EWMA), cumulative sum (CUSUM), and moving average (MA) charts (Ali et al 2019). The need for these different charts can be attributed to the search for a control charting technique that adequately satisfies objectives such as minimizing the probability of a false alarm, enhancing the ability to detect small shifts, and accounting for autocorrelation. However, the application of the aforementioned charts has primarily been confined to the univariate case in the industry due to their ease of use and interpretation. (Thomas, 2002)

2.2. Multivariate Control Chart

A control chart normally monitors one variable over time. Perhaps this variable is machine uptime, a product characteristic, or on-time delivery. There are times, however, when the simultaneous monitoring of two or more related variables is important. The group of control charts that do this is called multivariate control charts. The most familiar one of these is the Hotelling T^2 control chart or just the T^2 control chart. This control chart is introduced in this publication, (Bill, 2019), (Ali et al 2023). Hotelling T^2 statistic was the first statistic known to be used in a multivariate control chart. This statistic is used to measure the significance of the shifted distance from the out-of-control mean vector, to the nominal mean vector, with the assumption that the covariance matrix remains constant. (Hazlina, 2013)

Often in industrial settings, the overall quality of a product is not determined by a single characteristic in a process but is a function of many variables. For many years, separate univariate control charts have been used to m monitors the consistency of the quality of each variable in a multivariate process over time. The basic assumption associated with this technique is that each variable functions independently of the other variables in the process; however, this assumption is often invalid in practice. Therefore, the need to m monitor the overall process, which includes accounting for the correlation structure between the variables as well as controlling each variable's quality, presents a m much-needed improvement to simply using univariate techniques in a multivariate system.

Research into and implementation of multivariate control methods have become e more practical due to the advent of more powerful computing facilities. As computing technology continues to improve, the ease of data collection and manipulation also improves. With the abundance of data available today, managers are becoming aware of the necessity of utilizing as much information as possible in monitoring the quality of their processes. Studies have shown that the poor quality of a process may not always be due to one variable but may be attributed to several variables (Hotelling 1947; Jackson 1985).

Multivariate quality control procedures have been developed to make use of the correlation structure among the variables when determining if the process is in control. Some strengths of these approaches are that the overall probability of a false alarm can be accurately determined, the variables' correlation structure is reflected in the charts, and the number of control charts to monitor is reduced from one per quality characteristic to one for the entire process. Because of these important reasons, multivariate quality control has gained much attention while at the same time, it has created new concerns. One m major problem is detecting which variable(s) are responsible for out-of-control conditions. Another problem is the view of multivariate quality control charts as computationally complex and laborious to interpret. However, to provide the best possible decision-making tools using available information, multivariate quality control charts are necessary (Ali et al, 2018).

There are many situations where simultaneous monitoring is necessary to control two or more related quality characteristics and monitor whether these characteristics can be misleading. For these situations, specific tools should be used to detect, identify, and analyze the meaningful causes of variability in a process. Multivariate control charts represent one of these techniques being used to simultaneously control several characteristics that indicate the quality of a single production process. The most familiar monitoring and control procedure of a multivariate process is the Hotelling T^2 control chart, for monitoring the mean vector of the process. It is directly analogous to the univariate Shewhart X_ chart (Montgomery, 2009). Hotelling was the first researcher to know the weakness of the univariate statistical control charts in his pioneering paper. In the following decades, many contributors have established studies in the same field and extensive literature can be found, e.g.: Jackson (1985); Tracy, Young, and Mason (1992); Lowry and Montgomery (1995); Aparisi (1997); Nedumaran and Pignatiello Jr. (1999), Khoo et al. (2005); Champ and Farmer (2007), Bersimis et al. (2007), Frisen (2011), (Ali, 2017), (Ali and Saleh 2021) and (Ali and Jwana 2022).

Among the existing multivariate charts, the Hotelling T^2 control chart is the best known in the literature, and its applicability is most recommended for processes that have several quality characteristics. These characteristics are correlated and need to be monitored together. The T^2 test statistic is based on Equation (1), (Willems et al 2002)

$$T^2 = n(\bar{x} - \bar{\bar{x}})' S^{-1}(\bar{x} - \bar{\bar{x}}) \tag{1}$$

where $\bar{x}$ corresponds to the vector of means, $\bar{\bar{x}}$ and S represents the covariance matrix of the process. The application of the Hotelling T^2 multivariate control chart is done in two steps. In phase I, the limits are calculated using Equation (2) (Henning et al 2014).

$$UCL = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1} \tag{2}$$

where p is the number of variables, m is the number of samples, n is the sample size, and F equals Snedecor's F distribution with a degree of freedom for the numerator equal to α (equivalent to the rate of false positives), and for the denominator equal top, mn-m-p+1.

For phase II of the application of the multivariate chart, the equation of the upper control limit is given by the Equation (3) (Tracy et al 1992; Bersims et al 2007),

$$UCL = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1} \quad (3)$$

The lower control limit (LCL) for the two phases is equal to zero. For the use of multivariate control charts, it is also necessary to verify the assumptions of normality and independence. If the multivariate normal is not an appropriate model, there is very little literature available on alternative multivariate charting techniques (Bersimis et al 2007) such as multivariate non-parametric statistical process techniques (Chakraborti et al 2001) the autocorrelated multivariate process is also an area that must be further investigated.

3. Robust Method

Computes the robust covariance matrix estimate of a multivariate data set X, where X is an N-by-P matrix where each row is an observation and each column a variable (Ali and Saleh. 2022). Supported algorithms are:

I. The FASTMCD algorithm by Rousseuw and Leroy (FMCD).
II. The OGK algorithm by Maronna and Zamar.
III. The family of "concentration algorithms" by Olive and Hawkins (OH).

which will be explained in detail along with the proposed methods in the formation of the robust charts in Section 4

4. Proposed Chart

The research proposal is to create a multivariate quality control chart corresponding to the Shewhart (Hotelling) T^2-Chart, based on robust estimates of the variables averages and the covariance matrix, and for three methods, namely:

Proposed 1: depending on the algorithm Rousseuw and Leroy, the FAST-MCD (Minimum Covariance Determinant) method, the points drawn on the chart are calculated for the proposed chart and are first created (phase I). This method looks for h observations out of N (where N/2 < h <= N) whose classical covariance matrix has the lowest possible determinant. The estimate is then the covariance matrix of the h points defined above, multiplied by a consistency factor to obtain consistency at the multivariate normal distribution, and by a correction factor to correct for bias at small samples. On this basis, we obtain the estimates of the general average vector (MR) of the studied variables and the covariance matrix (SR) that is robust to outliers.

Where MR = (MRx_1, MRx_2, …, MRx_p), the robust mean vector represents for each variable (i = 1, 2, …, p), and robust covariance matrix (SR) is:

$$SR = \begin{pmatrix} sr_{x_1x_1} & sr_{x_1x_2} & \cdots & sr_{x_1x_p} \\ sr_{x_2x_1} & sr_{x_2x_2} & \cdots & sr_{x_2x_p} \\ \vdots & \vdots & \vdots & \vdots \\ sr_{x_px_1} & sr_{x_px_2} & \cdots & sr_{x_px_p} \end{pmatrix} \quad (4)$$

That is, the sample robust variances on the main diagonal of the matrix SR, and sample robust covariances on the off-diagonal of the matrix SR, therefore then computed as

$$T_R^2 = n(\bar{x} - MR)' SR^{-1}(\bar{x} - MR)) \quad (5)$$

Formula (5) is used to calculate the points drawn on the chart, while the upper control limit is as in Shewhart's chart on the tabular value of F (Kareem et al 2019).

Proposed 2: depending on the algorithm Maronna and Zamar, and using the Orthogonalized Gnanadesikan-Kettenring (OGK) estimate, the points drawn on the chart are calculated for the proposed chart and are first created (phase I). This estimate is a positive definite estimate of scatter starting from the Gnanadesikan and Kettering (GK) estimator, a pairwise robust scatter matrix that may be non-positive definite. The estimate uses a form of principal Components, called an orthogonalization iteration, on the pairwise scatter matrix, replacing its eigenvalues, which could be negative, by robust variances. This procedure can be iterated for improved results, and convergence is usually obtained after 2 or 3 iterations. On this basis, we obtain the estimates of the general average vector (MR) as in proposed 1 of the studied variables and the covariance matrix (SR) as in formula (4) that is robust to outliers. Therefore then T_R^2 is computed as formula (5), and is used to calculate the points drawn on the chart, while the upper control limit is as in Shewhart's chart on the tabular value of F.

Proposed 3: depending on the family of "concentration algorithms" by Olive and Hawkins (OH), and using the "concentration algorithm" techniques proposed by Olive and Hawkins, a family of fast, consistent, and highly outlier-resistant methods, the points drawn on the chart are calculated for the proposed chart is first created (phase I). The estimate is obtained by first generating trial estimates or starts, and then using the concentration technique from each trial fit to obtain attractors. By default, two attractors are used. The first attractor is the DGK (Devlin-Gnanadesikan-Kettering) attractor, where the start is the classical estimator (Shahla et al 2023). The second attractor is the Median Ball (MB) attractor, where the start used is Mu = median of data and Sigma is identity matrix by (p × p), i.e., the half set of data closest to a median of data in Euclidean distance. The MB attractor is used if the location estimator of the DGK attractor is outside of the median ball, and the attractor with the smallest determinant is used otherwise. The final mean estimate is the mean estimate of the chosen attractor, and the final covariance estimate is the covariance estimate of the chosen attractor, multiplied by a scaling factor to make the estimate consistent at the normal distribution. On this basis, we obtain the estimates of the general average vector (MR) as in proposed 1 of the studied variables and the covariance matrix (SR) as in formula (4) that is robust to outliers. Therefore then T_R^2 is computed as formula (5), and is used to calculate the points drawn on the chart, while the upper control limit is as in Shewhart's chart on the tabular value of F (Ali et al, 2018).

For all supported methods, a reweighting for efficiency step is performed. This step does not affect the robustness but improves the efficiency of the estimator.

5. Evaluation Criteria

For comparison between the classical and the proposed chart, the total and generalized variance can be used for the covariance matrix as follows:

$$Total\ Variance = trace\ (SR) \quad (6)$$

$$General\ Variance = \ |SR| \quad (7)$$

The lowest value for the total and generalized variance is the best (Ali et al 2023).

6: Application aspect

To compare the classical and the proposed charts in terms of efficiency and accuracy of the estimated multivariate mean vector and covariance matrix, the simulation study was done by simulating the multivariate quality control chart, then the application for the real data based on total and generalized variance. And by designing a program in MATLAB (version 2022a) dedicated to this purpose (Appendix).

6.1: Simulation study

Two samples (p = 2) were generated for the multivariate normal distribution (Appendix), with a correlation coefficient (0.70), and several observations (125) for (m = 25) subsamples and n = 5, for all subsamples, and outliers are randomly added to the generated data (normal distribution), (Ali, 2022). The first simulation experiment with the values of Mahalanobis distance is shown in Figure (1) for the classical method and three robust methods.

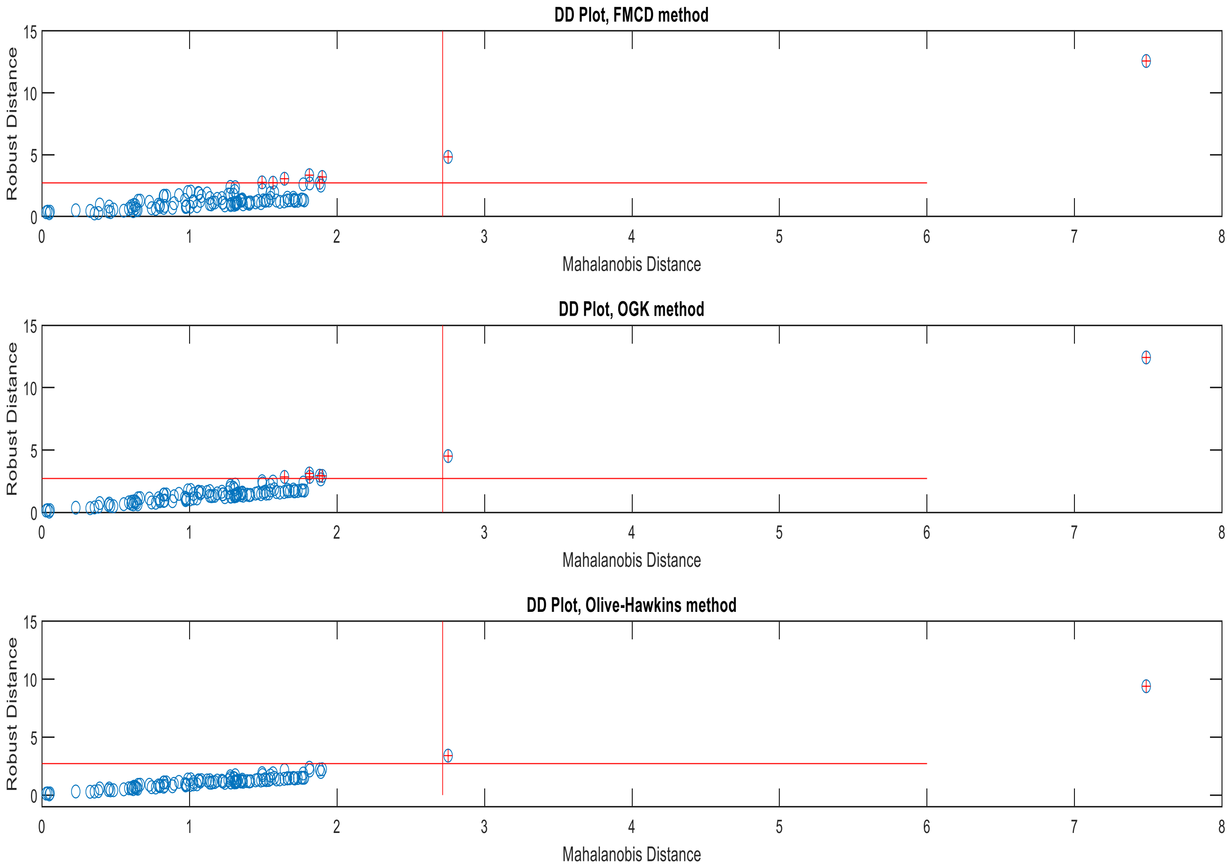

Figure 1. Mahalanobis distance for the classical method and three robust methods

Figure (1) shows that the Mahalanobis distance values for the classical method have two outliers, while there are (8, 7, and 2) as outliers (red stars) for the robust methods (FMCD, OGK, and OH), respectively. The classical T^2-Chart (Phase-I) is configured as in the figure (2):

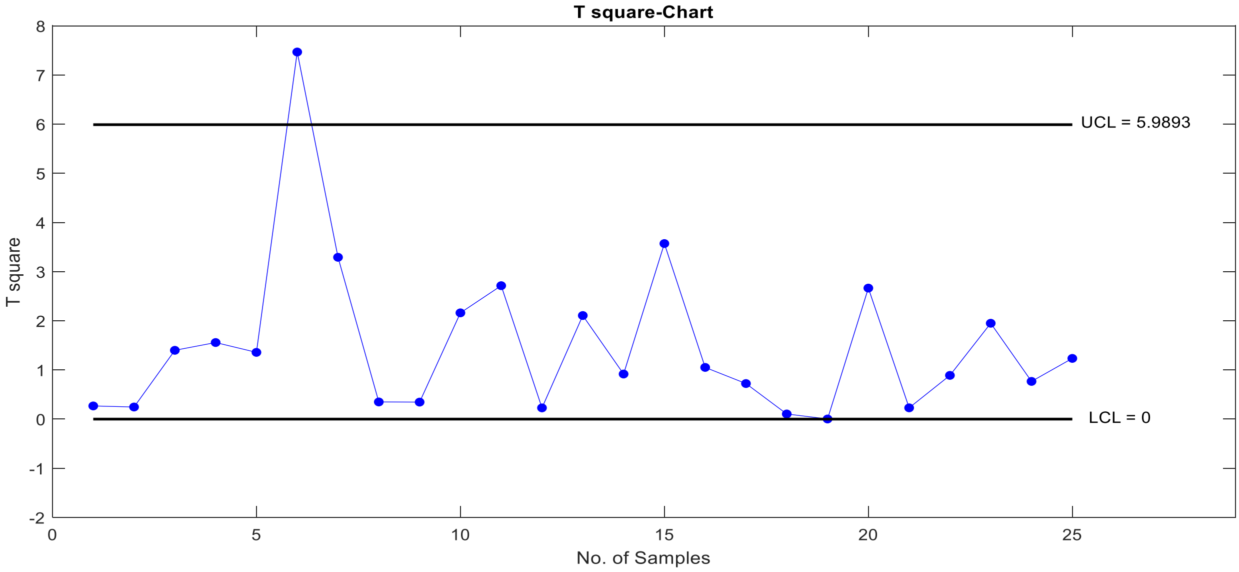

Figure 2. Classical T square-Chart

Figure (2) shows that there is one point outside the limits of control, so it will be deleted and a modified T2-Chart will be formed, as in Figure (3).

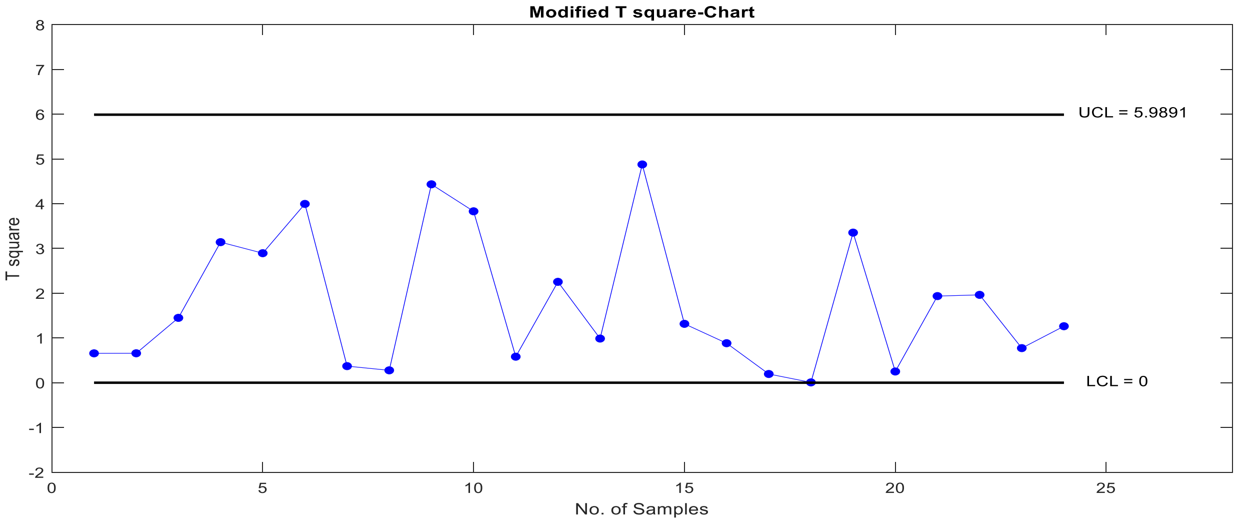

Figure 3. Modified Classical T square-Chart

The modified T^2-Chart shows that all the points drawn on the chart are within the control limits, so they can be relied upon and used in the future (Phase II).

The Proposed (and modified) T^2-Chart (Phase-I) is configured as in figure (4-9):

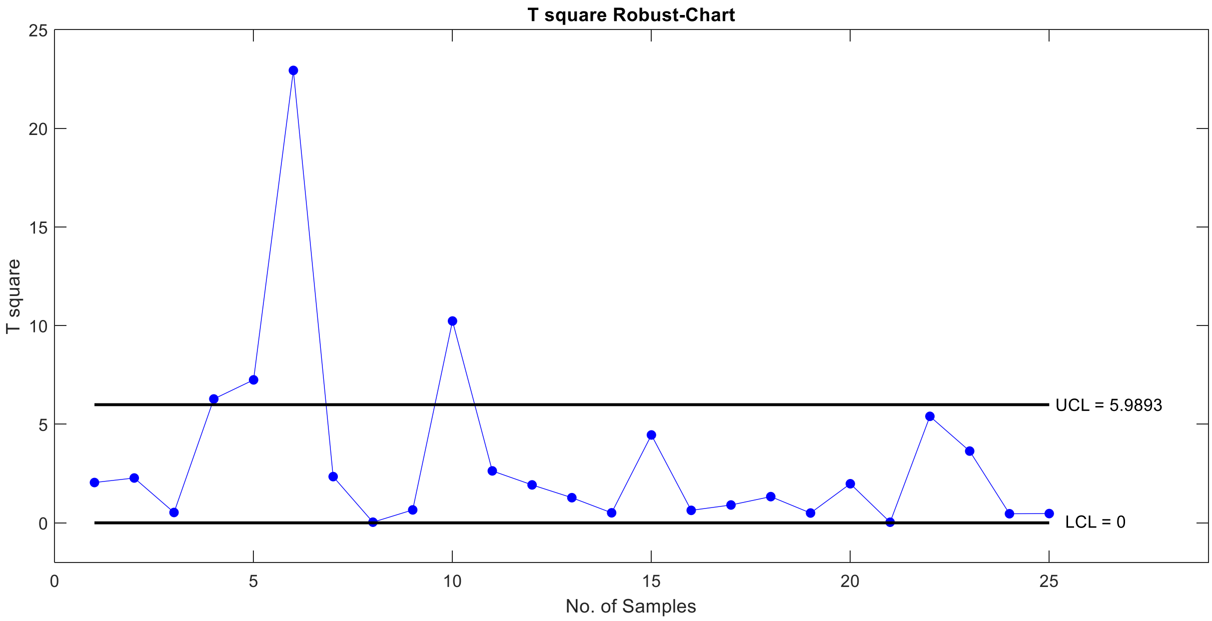

Figure 4. Proposed (1) T square Robust-Chart

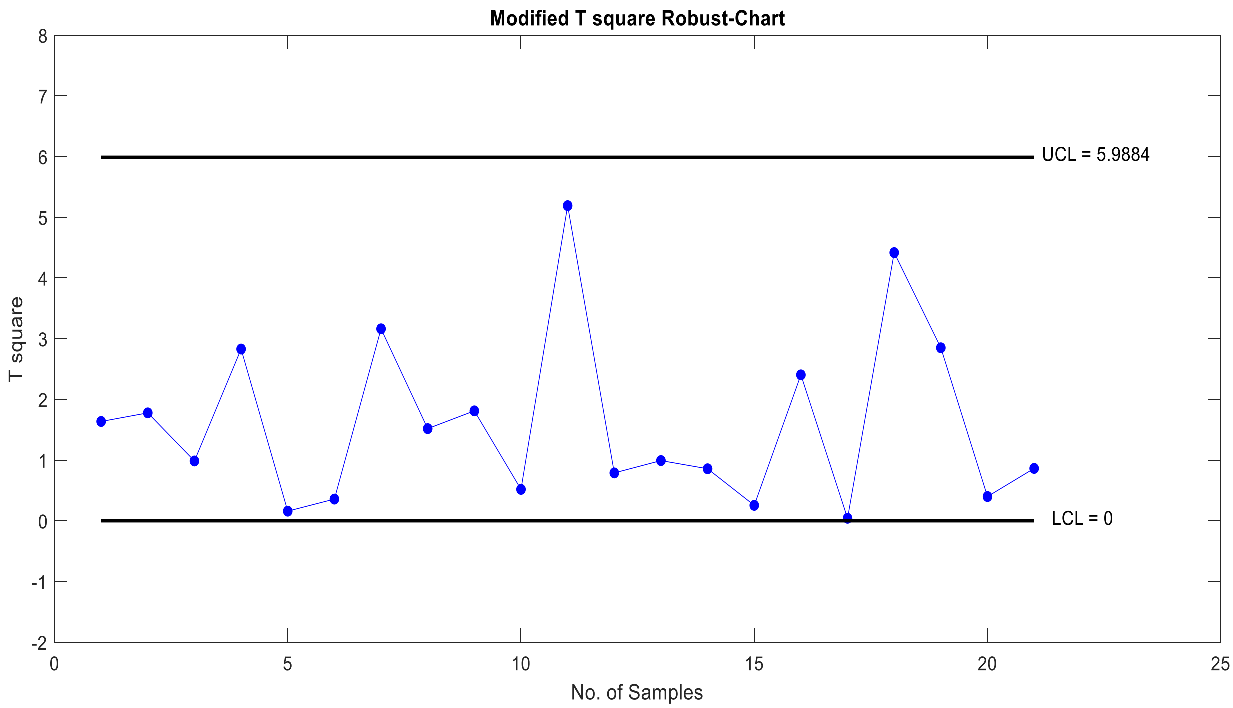

Figure 5. Modified Proposed (1) T square Robust-Chart

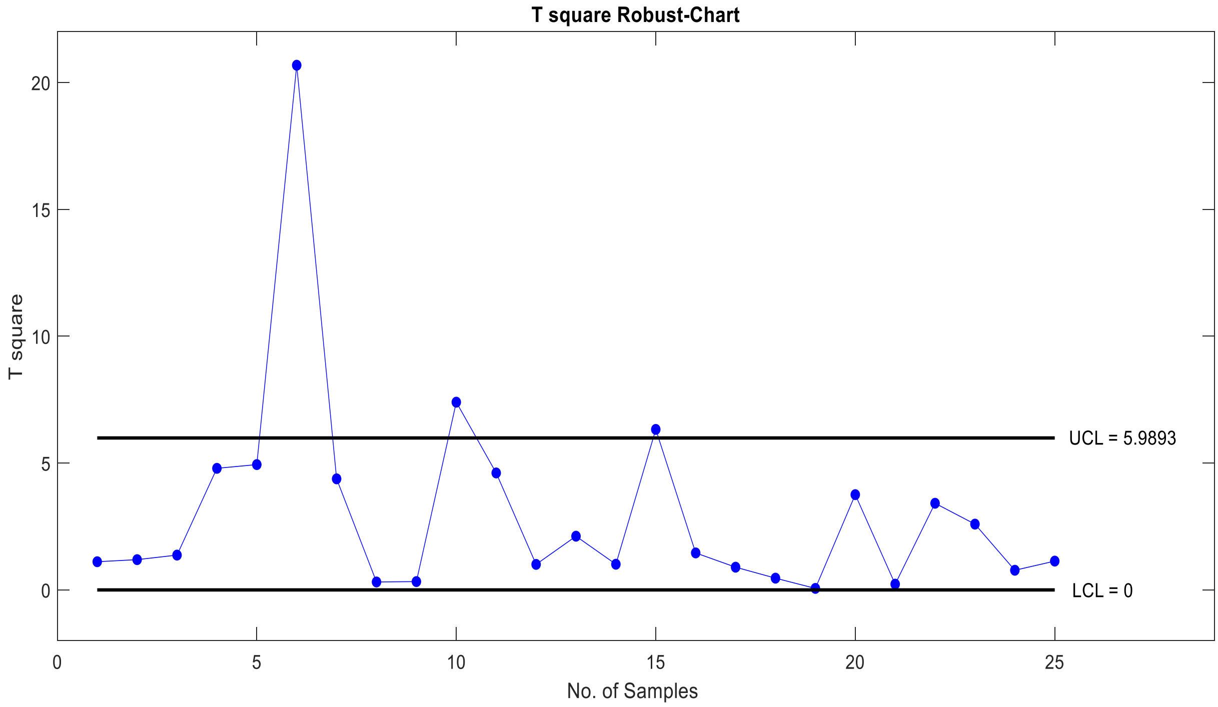

Figure 6. Proposed (2) T square Robust-Chart

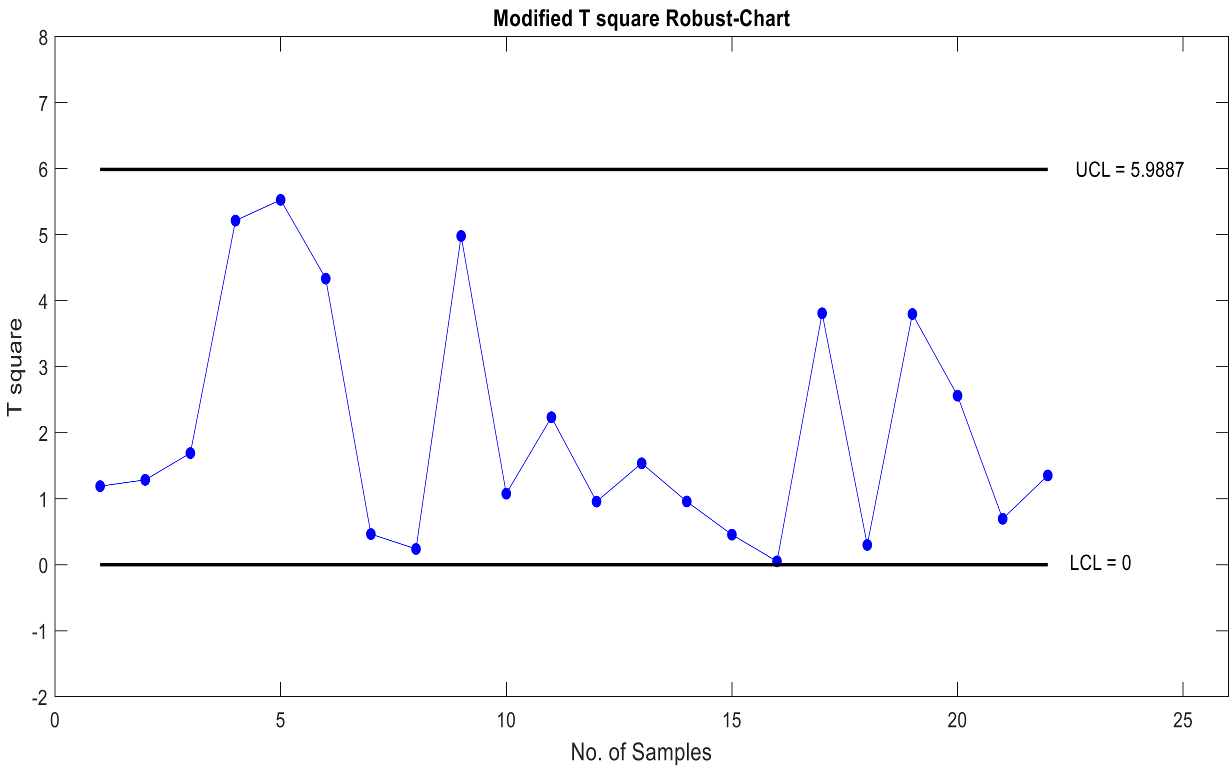

Figure 7. Modified Proposed (2) T square Robust-Chart

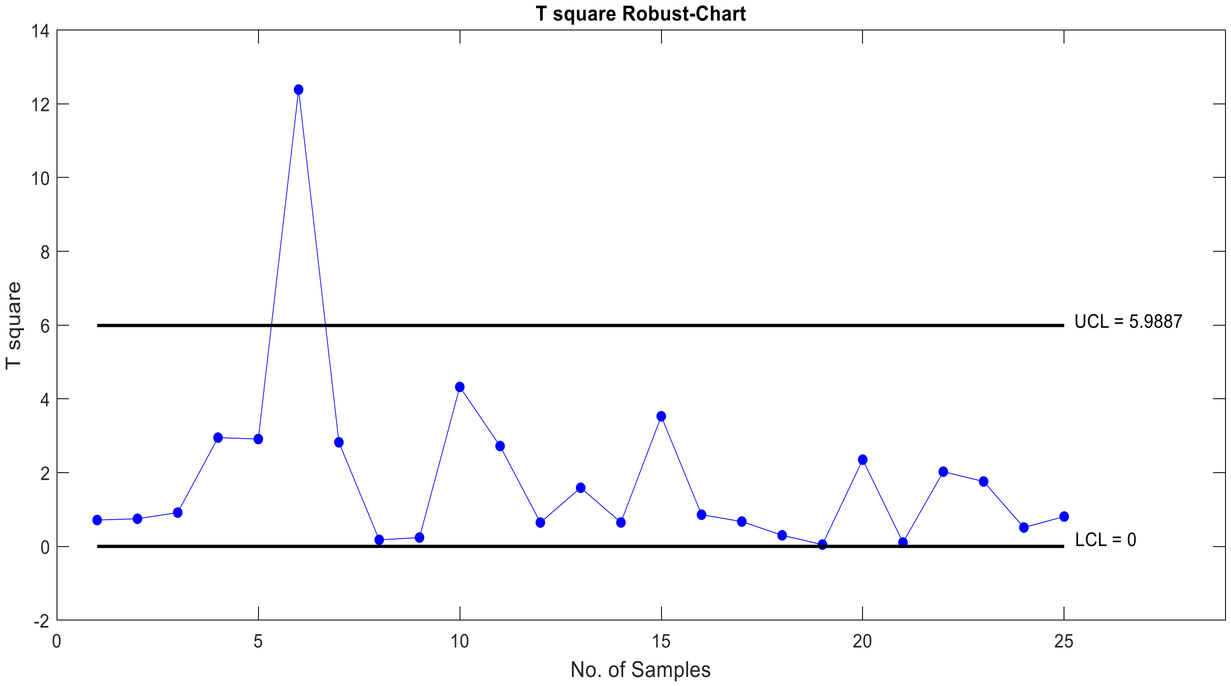

Figure 8. Proposed (3) T square Robust-Chart

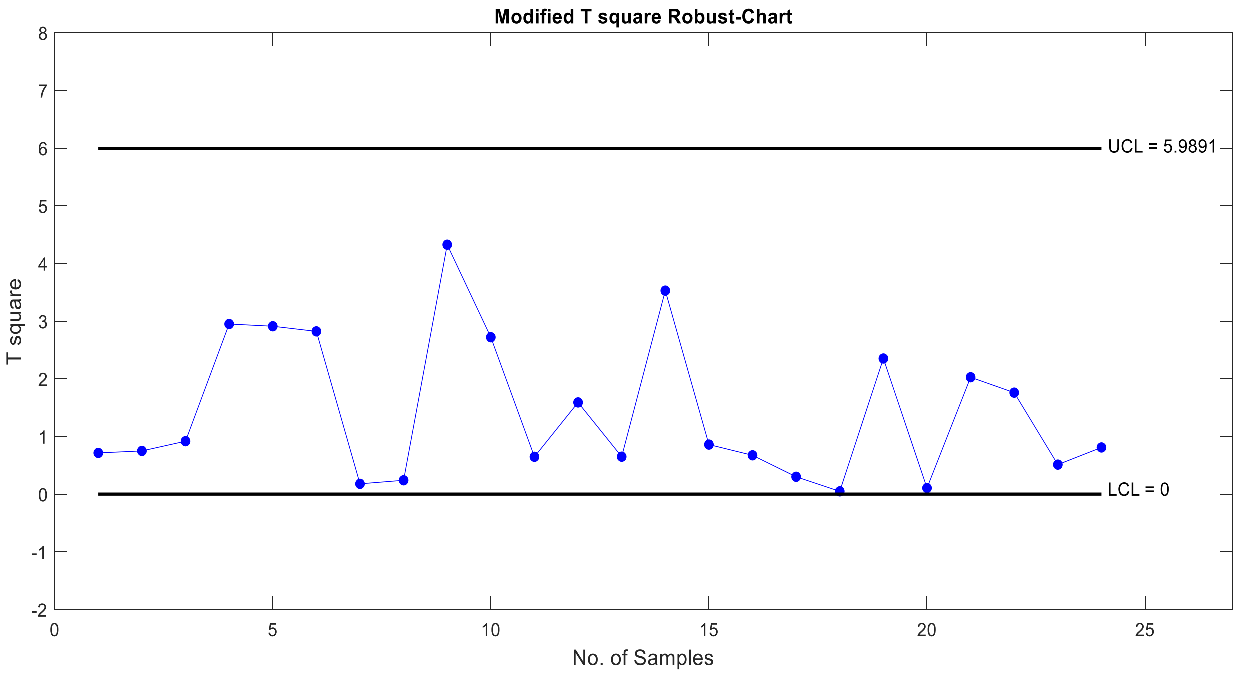

Figure 9. Modified Proposed (3) T square Robust-Chart

The modified proposed T^2-Charts show that all the points drawn on the chart are within the control limits, so they can be relied upon and used in the future (Phase II).

The simulation results of the first experiment for the classical and proposed charts are summarized in table (1):

Table 1. Results of the First Experiment

Chart	R	Generalized variance	Total variance	Average	UCL
Classical	0.6431	0.0062	0.2088	0.5440	5.9893

				0.4841	
Modified Classical	0.7631	0.0035	0.1822	0.5277 0.4802	5.9891
Proposed 1 (FMCD)	0.9070	0.0039	0.2954	0.5180 0.5144	5.9893
Modified Proposed 1 (FMCD)	0.8776	0.0031	0.2339	0.5053 0.4885	5.9884
Proposed 2 (OGK)	0.8392	**0.0023**	**0.1780**	0.5253 0.4888	5.9893
Modified Proposed 2 (OGK)	0.8670	**0.0021**	**0.1832**	0.5181 0.4781	5.9887
Proposed 3 (OH)	0.7946	0.0056	0.2473	0.5219 0.4890	5.9893
Modified Proposed 3 (OH)	0.7946	0.0056	0.2473	0.5219 0.4890	5.9891

Table (1) shows the classical and robust correlation coefficients for the primary and modified charts, the upper control limit, the mean vector, and the general and total variance. All the robust methods were better than the classical method, while the second proposed method (OGK) was the best based on the general and total variance. Also, the robust mean vector was lower than the classical mean vector due to the treatment of outliers.

The experiment was repeated (1000) times for sample sizes (5, and 10), 2 several variables, and the correlation coefficient (0.70), then the average of the general and total variance was calculated, and the results are summarized in Table (2):

Table 2. Average results of a thousand experiments (n = 5 and m = 25)

Chart	R	Generalized variance	Total variance	Average
Classical	0.5296	0.0099	0.2471	0.5350 0.5001
Proposed 1 (FMCD)	0.7755	0.0048	0.2293	0.4986 0.4992
Proposed 2 (OGK)	0.7332	**0.0031**	**0.1646**	0.5000 0.4984
Proposed 3 (OH)	0.7011	0.0065	0.2275	0.4998 0.4989

All the proposed charts were better than the classical chart when n = 5, because the averages of the general (0.0048, 0.0031, and 0.0065) and total (0.2293, 0.1646, and 0.2275) variance for the proposed charts were less than the classical chart (0.0099 and 0.2471), respectively. The robust chart (OGK) was better than the other robust charts. For the robust methods, the correlation coefficients were greater than the classical method, as well as the estimators of the robust mean vectors were less than the classical mean vector.

Table 3. Average results of a thousand experiments (n = 10 and m = 25)

Chart	R	Generalized variance	Total variance	Average
Classical	0.4171	0.0202	0.3725	0.5415 0.4994
Proposed 1 (FMCD)	0.7999	0.0045	0.2312	0.4992 0.4994
Proposed 2 (OGK)	0.7318	**0.0031**	**0.1648**	0.5002

				0.4989
Proposed 3 (OH)	0.6964	0.0066	0.2268	0.5001 0.4991

From Table (3), we note that all the proposed charts were better than the classical chart when n = 10, because the averages of the general (0.0045, 0.0031, and 0.0066) and total (0.2312, 0.1648, and 0.2268) variance for the proposed charts were less than the classical chart (0.0202 and 0.3725), respectively. Also, the robust chart (OGK) was better than the other robust charts. For the robust methods, the correlation coefficients were greater than the classical method, as well as the estimators of the robust mean vectors were less than the classical mean vector. Finally, when the sample size was increased, the classical chart was less accurate, while the proposed charts maintained their accuracy.

6.2. Real data

The real data contains various measured variables for about 200 automobiles from the 1970s and 1980s. We'll illustrate multivariate visualization using the values for fuel efficiency (in miles per gallon, MPG), acceleration (time from 0-60MPH in a sec), engine displacement (in cubic inches), and horsepower. We'll use the number of cylinders to group observations for (m = 20) subsamples and n = 10, for all subsamples. The real data with the values of Mahalanobis distance is shown in Figure (10) for the classical method and three robust methods.

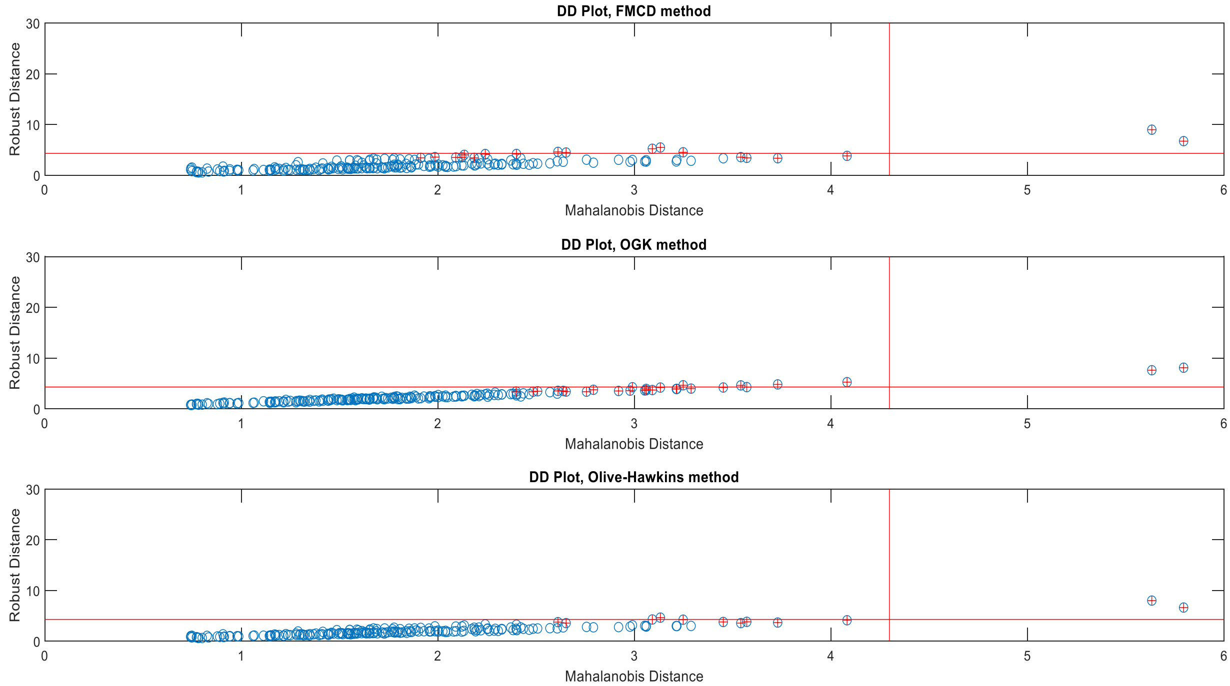

Figure 10. Mahalanobis distance for real data and three robust methods

Figure (10) shows that the Mahalanobis distance values for the classical method have several outliers (red stars), for the robust methods (FMCD, OGK, and OH). The classical T^2-Chart (Phase-I) is configured as in the figure (11):

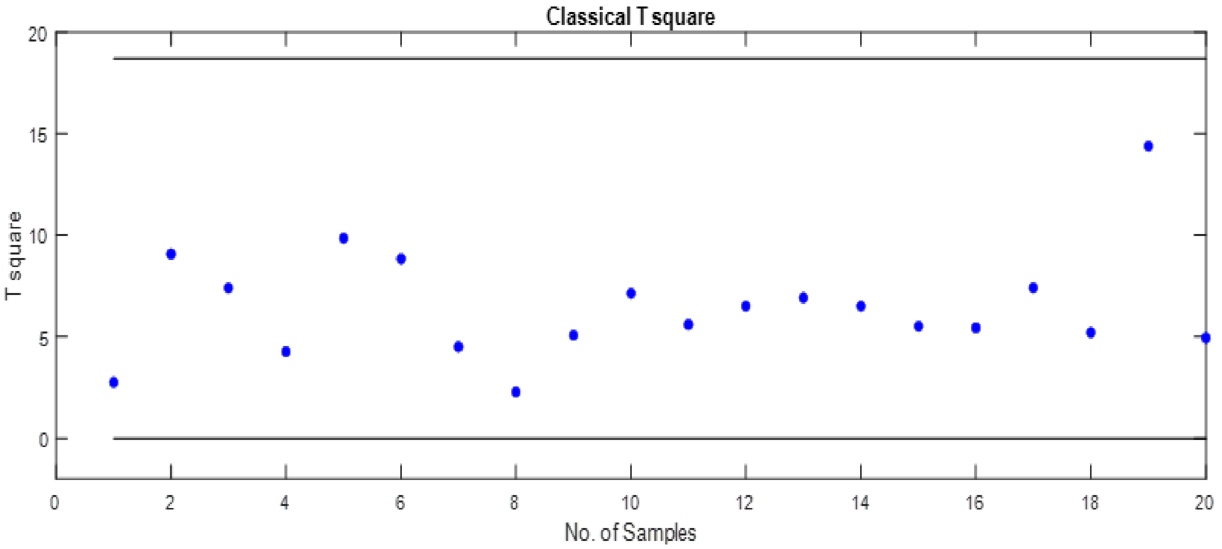

Figure 11. Classical T square-Chart for real data

The T^2-Chart shows that all the points drawn on the chart are within the control limits, so they can be relied upon and used in the future (Phase II). The Proposed (OGK) T^2-Chart (Phase-I) is configured as in the figure (12):

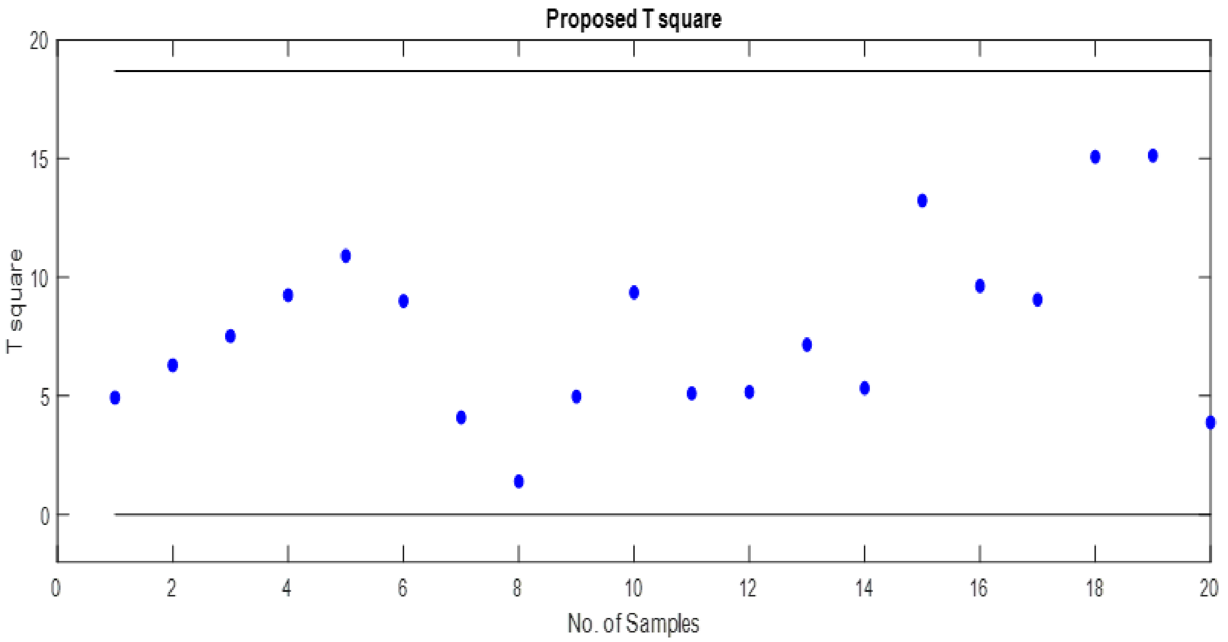

Figure 12. Proposed (2) T square Robust-Chart for real data

The Proposed (OGK) T^2-Chart shows that all the points drawn on the chart are within the control limits, so they can be relied upon and used in the future (Phase II). The real data results for the classical and proposed (OGK) chart are summarized in table (4):

Table 4. Results of real data

Chart	Generalized variance	Total variance	Average	UCL
Classical	60151000	9433.4	21.5150 15.8355 190.230 102.075	18.6948
Proposed 2 (OGK)	**39538000**	**8263.3**	22.4017 16.1523	18.6948

			173.586 95.3506	

Table (4) shows the proposed method (OGK) was the best based on the general and total variance. Also, the robust mean vector was lower than the classical mean vector due to the treatment of outliers.

7. Conclusion & Recommendations

Through the study of simulation and real data, the following main conclusions and recommendations were summarized:

7.1 Conclusions

1- All the proposed charts were better than the classical method.
2-The robust chart (OGK) was better than the other robust charts.
3-The robust correlation coefficients were greater than the classical correlation coefficient.
4- The estimators of the robust mean vectors were less than the classical mean vector.
5-when the sample size was increased, the classical chart was less accurate, while the proposed charts maintained their accuracy.

7.2 Recommendations

1. Using robust methods when outliers are present in construction the T^2-Chart, and especially the robust (OGK) method.
2. Conducting a prospective study on the use of robust methods in the construction of the S-Chart.
3. Conducting a prospective study on the use of multivariate wavelet in the construction of the T^2-Chart.
4. Conducting a prospective study on the use of multivariate wavelet with robust methods in the construction of the T^2-Chart and S-Chart.

8. References

- Ali, Taha Hussein & Awaz Shahab M. (2017) "Uses of Waveshrink in Detection and Treatment of Outlier Values in Linear Regression Analysis and Comparison with Some Robust Methods", Journal of Humanity Sciences 21.5:38-61.
- Ali, Taha Hussein and Jwana Rostam Qadir. (2022) "Using Wavelet Shrinkage in the Cox Proportional Hazards Regression model (simulation study)", Iraqi Journal of Statistical Sciences, 19, 1.17-29.
- Ali, Taha Hussein, and Dlshad Mahmood Saleh. (2021) "Comparison Between Wavelet Bayesian and Bayesian Estimators to Remedy Contamination in Linear Regression Model" PalArch's Journal of Archaeology of Egypt/Egyptology 18.10: 3388-3409.
- Ali, Taha Hussein, and Saleh, Dlshad Mahmood, (2022) "Proposed Hybrid Method for Wavelet Shrinkage with Robust Multiple Linear Regression Model: With Simulation Study" QALAAI ZANIST JOURNAL 7.1: 920-937.
- Ali, Taha Hussein, Heyam Abd Al-Majeed Hayawi, and Delshad Shaker Ismael Botani. (2023) "Estimation of the bandwidth parameter in Nadaraya-Watson kernel non-parametric regression based on universal threshold level." Communications in Statistics-Simulation and Computation 52.4: 1476-1489.
- Ali, Taha Hussein, Nazeera Sedeek Kareem, and Mohammad, Awaz Shahab (2019) "Construction robust simple linear regression profile Monitoring" Journal of Kirkuk University for Administrative and Economic Sciences, 9.1.: 242-257.
- Ali, Taha Hussein, Rahim, Alan Ghafur, and Saleh, dlshad Mahmood. (2018) "Construction of Bivariate F-Control Chart with Application" EURASIAN JOURNAL OF SCIENCE AND ENGINEERING (EAJSE), 4.2: 116-133.
- Ali, Taha Hussein, Saman Hussein Mahmood, and Awat Sirdar Wahdi. (2022) "Using Proposed Hybrid method for neural networks and wavelet to estimate time series model." Tikrit Journal of Administration and Economics Sciences 18.57 part 3.

- Ali, Taha Hussein. (2022) "Modification of the adaptive Nadaraya-Watson kernel method for nonparametric regression (simulation study)." Communications in Statistics-Simulation and Computation 51.2: 391-403.
- Ali, Taha Hussein; Esraa Awni Haydier. (2016) "Using Wavelet in constructing some of Average Charts for Quality control with application on Cubic Concrete in Erbil", Polytechnic Journal, 6.2: 171-209.
- Ali, Taha Hussein; Saleh, Dlshad Mahmood; Rahim, Alan Ghafur. (2017) "Comparison between the median and average charts using applied data representing pressing power of ceramic tiles and power of pipe concrete", Journal of Humanity Sciences 21.3: 141-149.
- Ali, Taha Hussien, (2017), "Using Proposed Nonparametric Regression Models for Clustered Data (A simulation study)." Journal of Humanity Sciences, 29.2: 78-87.
- Bersimis, S.; Psarakis, S.; Panaretos, J., (2007) "Multivariate Statistical Process Control Charts": An Overview. Quality and Reliability Engineering International, v. 23, n. 5, p. 17-543.
- Bill McNeese (2019) "Multivariate Control Charts: The Hotelling T2 Control Chart" BPI Consulting, LLC www.spcforexcel.com.
- Chakrabortl, S.; Van Der Laan, P.; Bakir, S. T. (2001) "Nonparametric control charts: An overview and some results" . Journal of Quality Technology, v. 33, n. 3, p. 304-315.
- Charu C. Aggarwal (2017) "Outlier Analysis. second Edition", New York, Springer International Publishing.
- Deming, W. E. (2000). Out of the crisis. Cambridge, MA: MIT Press.
- Farcomeni A., Greco L. (2021) "Robust Methods for Data Reduction" first edition Published March 31, by Chapman & Hall.
- Hazlina Ali, Sharifah Soaad Syed Yahaya, and Zurni Omar (2013 " Robust Hotelling *T*2 Control Chart with Consistent Minimum Vector Variance". College of Art and Sciences, University Utara Malaysia, 06010 Sintok, Kedah, Malaysia.
- Henning E., Mala M.T., Walter O., Konrath A., Alives C. (2014) "Application of hoteling T2 Control Chart for a machining Prpcess of the inside diameter of a steel Cylinder Bauru", Ano 9, n02, abr-jun-p. 155-167.
- Hotelling, H. (1947) "Multivariate Quality Control Illustrated by Air Testing of Sample Bombsights". In: Eisenhart, C., Hastay, M.W. and Wallis, W.A., Eds., Techniques of Statistical Analysis, McGraw Hill, New York, 111-184.
- Jackson, J. E (1985). "Multivariate quality control. Communications in Statistics. Theory and Methods", v. 14, n. 11, 2657-2688.
- Kareem, Nazeera Sedeek, Taha Hussein Ali, and Awaz shahab M, (2020) "De-noise data by using Multivariate Wavelets in the Path analysis with application", Kirkuk University Journal of Administrative and Economic Sciences, 10.1: 268-294.
- Kovach, J. (2007). "Designing Effective Six Sigma Experiments for Service Process Improvement Projects". International Journal of Six Sigma and Competitive Advantage, 3(1), 72–90.
- Montgomery, Douglas, C. (2009) “Introduction to Statistical Quality Control” Sixth Edition, Arizona state university, John Wiley & Sons, Inc. Printed in the United States of America.
- Shahla Hani Ali, Heyam A.A. Hayawi, Nazeera Sedeek K., and Taha Hussein Ali, (2023) "Predicting the Consumer price index and inflation average for the Kurdistan Region of Iraq using a dynamic model of neural networks with time series", The 7th International Conference of Union if Arab Statistician-Cairo, Egypt 8-9/3/2023:137-147.
- Thomas J. Brian (2002) " Interpretation of Multivariate Control Chart Signals "A Dissertation for degree of Doctor, University of Alabama, Tuscaloosa. Alabama.
- TRACY, N. D.; YOUNG J. C.; MASON, R. L. (1992) “Multivariate Control Charts for Individual Observations”. Journal of Quality Technology, v. 24, n. 2, p. 88-95.
- Willems G., Pison G., Rousseeuw P. J., and Van Aelst S. (2002) "A Robust Hotelling test" University of Antwerp, Metrika (125-138) Springer-Verlag.

Appendix

```
clc
clear all
rng default; n=5;m=25;K=2;J=125; rho = [1,0.7;0.7,1]; u = copularnd('Gaussian',rho,J);
noise = randperm(J,3); u(noise,1) = u(noise,1)*4; [Sfmcd, Mfmcd, dfmcd, Outfmcd] = robustcov(u);
[Sogk, Mogk, dogk, Outogk] = robustcov(u,'Method','ogk'); [Soh, Moh, doh, Outoh] =
robustcov(u,'Method','olivehawkins'); d_classical = pdist2(u, mean(u),'mahal'); p = size(u,2);
chi2quantile = sqrt(chi2inv(0.975,p)); tiledlayout(2,2); nexttile; plot(d_classical, dfmcd, 'o')
line([chi2quantile, chi2quantile], [0, 30], 'color', 'r'); line([0, 6], [chi2quantile, chi2quantile], 'color', 'r')
hold on; plot(d_classical(Outfmcd), dfmcd(Outfmcd), 'r+'); xlabel('Mahalanobis Distance')
ylabel('Robust Distance'); title('DD Plot, FMCD method'); hold off; nexttile; plot(d_classical, dogk, 'o')
line([chi2quantile, chi2quantile], [0, 30], 'color', 'r'); line([0, 6], [chi2quantile, chi2quantile], 'color', 'r')
hold on; plot(d_classical(Outogk), dogk(Outogk), 'r+'); xlabel('Mahalanobis Distance')
ylabel('Robust Distance'); title('DD Plot, OGK method'); hold off; nexttile; plot(d_classical, doh, 'o')
```

```
line([chi2quantile, chi2quantile], [0, 30], 'color', 'r'); line([0, 6], [chi2quantile, chi2quantile], 'color', 'r')
hold on; plot(d_classical(Outoh), doh(Outoh), 'r+'); xlabel('Mahalanobis Distance')
ylabel('Robust Distance'); title('DD Plot, Olive-Hawkins method'); hold off
number1 =sum(Outfmcd); number2 =sum(Outogk); number3 =sum(Outoh);
% quality control
x1=u(:,1);x2=u(:,2);x=[x1,x2]; [S M]=robustcov(x,'Method','fmcd'); r=corr(x);s=1; for i=1:m
   m1(i,:)=mean(x(s:s+4,:));     s=s+5;   end
% PROPOSED CHART
for i=1:m ; d(:,i)=m1(i,:)-M; T(:,i)=n*d(:,i)'*inv(S)*d(:,i); end; T;
U=1:m;alfa=.05;df1=K;df2=m*n-m-K+1; tabF=finv(1-alfa,df1,df2);
for i=1:m
UCL(i)=((K*(m-1)*(n-1))/(m*n-m-K+1))*tabF; LCL(i)=0;
end
plot(U,T,'bo',U,UCL,'-',U,LCL,'-')
R=S(1,2)/(sqrt(S(1,1)*S(2,2))), D=det(S), total=trace(S), AVERAG=M, UCL=UCL(1)
```

بنیاتنانی هێلکاری بەهێز T_R^2 و بەراوردکردنی لەگەڵ هێلکاری هۆتێلینگ T^2

بیخال صمد صدیق1
بەشی ئامارو زانیارییەکان ، کۆلێژی
بەرێوەبردن وئابووری،
زانکۆی سەلاحەددین-هەولێر
bikhal.sedeeq@su.edu.krd

زينب عبدالله محمد
بەشی ئامارو زانیارییەکان ، کۆلێژی
بەرێوەبردن وئابووری،
زانکۆی سەلاحەددین-هەولێر
zainab.muhammad@su.edu.krd

اسراء مۆیدعلي
بەشی ئامارو زانیارییەکان ، کۆلێژی
بەرێوەبردن وئابووری،
زانکۆی سەلاحەددین-هەولێر
israa.ali@su.edu.krd

طه حسين علي
بەشی ئامارو زانیارییەکان ، کۆلێژی
بەرێوەبردن وئابووری،
زانکۆی سەلاحەددین-هەولێر
taha.ali@su.edu.krd

پوختە

لەم توێژینەوەیەدا، پێشنیار کراوە بۆ بنیاتنانی هێڵکاریەکی نوێی فرەگۆڕاوی T_R^2 بەرامبەر هێلکاری T^2 بەهێزە دژی نرخی دەرەکی بە بەکارهێنانی سێ رێگا وەك (Rousseuw and Leroy) و (Maronna and Zamar) و (The family of "concentration algorithms" by Olive and Hawkins) پاشان بەراوردکردن لە نێوان رێگاکانی پێشنیارکراو کلاسیکی توێژەر شیوارت وە پشت بەستن بە جیاوازی گشتی (کۆی یەکەی تیرە سەرەکیەکان بۆ جیاوازیی ماتریکس) و جیاوازیی گشتی (جیاوازی ماتریکسی سنوردالرکراو بۆ بەدەست هێنانیی کاراترین هێلکاری بەهێز (رۆبەست) دژی نرخی دەرەرکی لە رێگای لاساییکردنەوە و داتاي راستەقینە بە بەکارهێنانی بەرنامەیەك لە زمانی ماتلاب کە بۆ ئەم مەبەستە دیزاین کراوە. لێکۆڵینەوەکە گەیشتە ئەو ئەنجامەی کە هێلکاری پێشنیارکراوەکان چارەسەری کێشەی کاریگەری نرخی دەرەکی دەکەن و کاراترن لە شێوازی کلاسیك.

ووشە سەرەکییەکان: هێلکاری دڵنیای جۆری فرەگۆڕاو، بەهێز، نرخی دەرەکی هێلکاریT^2.

تكوين لوحة T_R^2 الحصينة ومقارنتها مع لوحة هوتلنك - T^2

بيخال صمد صديق
قسم الاحصاء والمعلوماتية ،كلية
الادارة والاقتصاد ،
جامعة صلاح الدين – أربيل
bikhal.sedeeq@su.edu.krd

زينب عبدالله محمد
قسم الاحصاء والمعلوماتية ،كلية
الادارة والاقتصاد ،
جامعة صلاح الدين – أربيل
zainab.muhammad@su.edu.krd

اسراء مؤيدعلي
قسم الاحصاء والمعلوماتية ،كلية
الادارة والاقتصاد ،
جامعة صلاح الدين – أربيل
israa.ali@su.edu.krd

طه حسين علي
قسم الاحصاء والمعلوماتية ،كلية
الادارة والاقتصاد ،
جامعة صلاح الدين – أربيل
taha.ali@su.edu.krd

ملخص

تم في هذا البحث اقتراح تكوين لوحات جديدة لمتعدد المتغيرات T_R^2 مقابلة للوحة - T^2 حصينة ضد القيم الشاذة باستخدام ثلاث طرائق وهي خوارزمية (Rousseuw and Leroy)، (Maronna and Zamar) و (The family of "concentration algorithms" by Olive and Hawkins). ثم المقارنة بين الطرائق المقترحة والتقليدية للباحث شيوارت اعتماداً على التباين الكلي (مجموع عناصر القطر الرئيسي لمصفوفة التباين) والتباين العام (محدد مصفوفة التباين) للحصول على أكفأ لوحات حصينة ضد القيم الشاذة من خلال المحاكاة والبيانات الحقيقية وباستخدام برنامج بلغة ماتلاب مصمم لهذا الغرض. وتوصلت الدراسة إلى أن اللوحات المقترحة عالجت مشكلة تأثير القيم الشاذة وذات كفاءة أكبر من الطريقة التقليدية.

الكلمات المفتاحية: لوحات السيطرة النوعية متعددة المتغيرات، الحصينة، القيم الشاذة، لوحةT^2.