

A Comparison Between the MEWMA and Mahalanobis Distance Control Chart

Kawa M. Jamal Rashid
University of Sulamani
College Administration and Economics
Informatics Statistics Department
Email: kawa.rashid@univsul.edu.iq

Article Info

Volume 6 – Issue 2-
December 2021

DOI:
10.24017/science.2021.2.9

Article history:

Received 17/8/2021

Accepted 15/9/2021

Keywords:

Statistical Process Control (Exponentially Weighted Moving Average-EWMA and Multi Exponentially Weighted Moving Average (MEWMA) chart and Mahalanobis distance

ABSTRACT

Statistical Process Control (SPC) is approaching that uses statistical techniques to monitor the process. Quality control methods are used widely in charts strategy. A traditional variable control chart includes three lines: The Upper and Lower Control Limit (UCL LCL), and The Centre Line (CL), all of which are represented by numeric values. A control chart illustrates the centerline of the average value of the quality feature under investigation. Depending on numeric observation values, a process is either "in control" or "out of control." Considering there are no questions regarding the observations and their values during the production process. However, when these observations include human judgments, assessments, and choices, a continuous random variable (x_i) of a manufacturing process should be made up of the variable. The multivariate Exponential Weighted Moving Average (MEWMA) and Mahalanobis distance control chart techniques are used when more than one variable is involved. Today quality control has become one of the most essential techniques for studying all variables to control production or consumption decision factors. Quality control's primary aim is to guarantee that the goods, services, or processes delivered satisfy particular standards and are reliable and satisfactory. The factor's direct approach is based on the product's quality. This study shows that the MD-Distance control chart has a good result. This paper aims to compare the MEWMA and Mahalanobis distance chart for three compounds for water drinking production in the ALA -Factory in Sulaimani province of Iraq.

Copyright © 2021 Kurdistan Journal of Applied Research.
All rights reserved.

1. INTRODUCTION

Quality control methods are one of the most critical technologies to study all variables to control production. The primary strategy of quality control is to ensure that the products are in a high quality, processes provided meet consumers and are employed to ensure levels of quality in a product.

Quality control problems in the industry may involve more than a single quality characteristic. Walter A Shewhart (1924) was invented the quality control chart, which has formed the basis of extensive work in multivariate quality control; he was the first to recognize the need for considering quality control problems as multivariate.

During the period of 1930-1940 Hotelling performed a great deal of work on multivariate statistical control procedures. The Hotelling's T^2 control chart is one of the most widely used tools in multivariate statistical process control.

Continuous multivariate approaches include three important properties, multivariate Statistical Process Control is a collection of advanced techniques for monitoring and controlling and operational

(1) Is the procedure under your control?

(2) Type I errors must be specified.

(3) They must take into account the variables' interrelationships. [1][2][5][6].

Exponentially weighted moving average (EWMA) chart (Roberts, 1959) and cumulative sum CUSUM chart (Page, 1954) are often used for detecting shifts in a sequence of independent normal observations with common variance coming from a particular process.

The EWMA chart relies on the specification of a target value and a known or reliable standard deviation estimate. For this reason, the moving average chart is best used after process control has been established [6][8]. The exponentially weighted moving average (EWMA) control chart is a good alternative than Shewhart control chart.

In the Shewhart control chart when we are interested in detecting small shifts. The performance of the EWMA control chart is approximately equivalent to that of the cumulative sum control chart, and in some ways, it is easier to set up and operate. As with the CUSUM, the EWMA is typically used with individual observations [2][3][4]

If the observations X_i are independent random variables with variance S^2 , then the variance of Z_i is:

$$\sigma_{z_i}^2 = \sigma^2 \left(\frac{\lambda}{2 - \lambda} \right) [1 - (1 - \lambda)^{2i}] \quad (1)$$

2. THE MULTIVARIATE NORMAL DISTRIBUTION

The multivariate normal distribution model extends the univariate normal distribution model to fit multivariate observations. It is an essential direction in statistics that analyzes the relationships between more than one variable and analyzes the dependence between variables and between groups of variables.

Generally, we use the normal distribution to describe the behavior of a continuous quality characteristic.

The univariate normal probability density function is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2} \quad -\infty < x < \infty \quad (2)$$

The mean of the normal distribution is μ , and the variance is σ^2 . Note that (apart from the minus sign) the term in the exponent of the normal distribution can be written as follows:

$$(x - \mu)' \sigma^2 (x - \mu) \quad (3)$$

The measures of the squared standardized distance from X to the mean μ , whereby the term "standardized". The distance is expressed in standard deviation units.

The vector $x' = [x_1, x_2, \dots, x_p]$. Be a p -component of variables given by x_1, x_2, \dots, x_p , approach can be used in the multivariate normal distribution case.

Let $\mu' = [\mu_1, \mu_2, \dots, \mu_p]$

The variances and covariance's of the random variables in x be contained in a $(p \times p)$ covariance matrix S , and let the vector of the means of the x s be the vector of the means the x 's. The x 's variances are the major diagonal elements of Σ . The covariances are the off-diagonal elements. Now the normalized or standardized distance is squared [5][7].

From x to μ is:

$$(x - \mu)' \Sigma^{-1} (x - \mu) \quad (4)$$

In multivariate analysis, Mahalanobis distance (MD) has been a fundamental statistic, It was introduced by a famous Indian statistician Prof. P. C. Mahalanobis in 1936. It has been applied by researchers in several different areas [6][8][10].

The Mahalanobis distance (MD) measured the distance of unknown observations from the reference points in multivariate space or measuring the distance between vectors about different practical uses, such as the difference between pairwise individuals, comparing the similarity of observations. It's a multivariate equivalent of the Euclidean distance. One of the main reasons for using the MD is that it is very sensitive to intervariable changes in the reference data, the idea of measuring how many standard deviations from the mean. [6][9]

The Mahalanobis distance formula is as:

$$MD_j = D_j^2 = Z'_{ij} C^{-1} Z_{ij} \quad (5)$$

$$Z_{ij} = (z_{1j}, z_{2j}, \dots, z_{ij}) \quad (6)$$

$j = 1$ to n

$i = 1$ to k

$$Z_{ij} = \frac{X_{ij} - m_i}{S_i} \quad (7)$$

$$D^2 = (x - m)' C^{-1} (x - m) \quad (8)$$

Where:

D^2 : Is the MD

X : Is the vector of observations

m : Is the vector of mean

C^{-1} : Is the inverse covariance matrix

Finally, the Euclidean distance is calculated.

The three procedures listed above are intended to address the issues with Euclidean distance that we discussed before. So, how do you understand the formula above?

Let's take the $(x - m)' C^{-1} (x - m)$

Is essentially the vector's distance from the mean. The covariance matrix is then divided by this (or multiply by the inverse of the covariance matrix)

Is What is the effect of dividing by the covariance?

If the variables in your data set are strongly correlated, then the covariance will be high. Dividing by a significant covariance will effectively reduce the distance.

Likewise, if the X 's are not correlated, then the covariance is not high, and the distance is not reduced much.

So effectively addresses both the problems of scale and the vector's distance from the mean and the correlation of. The covariance matrix is then divided by this (or multiply by the inverse of the covariance matrix) [5][7].

Let X_i , $i = 1, \dots, n$ be random vectors with p components; its mean be $E(X_i) = \mu$ and covariance matrix be

$Cov(X_i) = \Sigma$, the MD space y_i is generated by

$$y_i = \Sigma^{-1/2} (x_i - \mu) \quad i = 1, 2, \dots, n \quad (9)$$

3. HOTELLING'S T² DISTRIBUTION

Some of the multivariate procedures for control charts are based heavily on Hotelling's T² distribution introduced by an American mathematical statistician and an influential economic Harold Hotelling's father theorist T² distribution (1947).

The basis of several multivariate control charts or a multivariate sampling distribution, the T² control chart can be used with data in subgroups or data that are individual observations to show how to construct a T² control chart[1][5] .

$$t^2 = \frac{(\bar{X} - \mu_0)^2}{s^2/n} \quad (10)$$

$$(x - \mu)'S^{-1}(x - \mu) \quad (11)$$

Where x is the sample mean vector and S is the sample covariance vector.

The S matrix is a little more challenging to find. It is found from the vector of moving differences for each variable. For each variable, Vi is found where:

$$V_i = X_{i+1} - X_i$$

This is done for both variables and the vector V contains the results for both variables:

$$V = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_{m-1} \end{bmatrix}$$

Where a m are number of samples. And V has a column for each variable and (m - 1) rows.

The S can be found using the following:

$$S = \frac{1}{2} \frac{V'V}{(m-1)} \quad (12)$$

Where V' is the transpose of V

Then T² is given by:

$$T^2 = (x - x')'S^{-1}(x - x)$$

For the T² control chart, we need to define or calculate is the UCL. And, of course, The upper control limit based on individual observations is given by the following:[7][8][11]

$$UCL = \frac{(m-1)^2}{m} \beta_{\alpha, \frac{p}{2}, (q-p-1)/2} \quad (13)$$

Where m = number of samples, p = number of variables, β = Beta distribution, α = the

confidence level and $q = \frac{2(m-1)^2}{3m-4}$ (14)

It is well known that for a random sample of size n from a univariate normal population, $X \sim N(\mu, \sigma^2)$, if we compute the sample mean \bar{X} And sample variance S^2 ,

Then it $t = (\bar{X} - \mu) / (\frac{s}{\sqrt{n}})$ Will have a t distribution with (n - 1) degrees of freedom[10][11]

Suppose that $X_1, X_2, X_3, \dots, X_n$, are jointly distributed with a multivariate with a normal distribution, Let the mean be as $\mu_1, \mu_2, \mu_3, \dots, \mu_n$ and variances as $\sigma_1^2, \sigma_2^2, \sigma_3^2, \dots, \sigma_{n1}^2$ For $n > 1$

The Characteristics may be independent and the covariance of X_i and X_j is given as

$$\sigma_{ij} = E(X_i - \mu_i)(X_j - \mu_j) \quad (15)$$

And the variance of X_i is:

$$\sigma_{ij} = E(X_i - \mu_i)(X_j - \mu_j) = E[(X_i - \mu_i)^2] = \sigma_i^2 \quad (16)$$

And the random variables can be shown as a vector:

$$X_k = \begin{bmatrix} X_{1k} \\ X_{2k} \\ \vdots \\ X_{pk} \end{bmatrix} \quad k = 1, 2, 3, \dots, m \quad j = 1, 2, 3, \dots, p \quad j \neq h \quad (17)$$

And the vector of E(X) value of X is:

$$\hat{\mu} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix}$$

Where

$$\bar{X}_j = \frac{\sum_{k=1}^m X_{jk}}{m}$$

and

$$S_j^2 = \frac{\sum_{k=1}^m (X_{jk} - \bar{X}_j)^2}{m-1} \quad \begin{matrix} j = 1, 2, 3, \dots, p \\ j \neq h \end{matrix} \quad (18)$$

Therefore we can calculate the value of Hoteling T^2

If we have two characteristics, X_1 and X_2 with a joint distribute with sample mean, variance and covariance between two variables, under these conditions the formula of Hoteling T^2 as:

$$T^2 = \frac{n}{S_1^2 S_2^2 - S_{12}^2} [S_2^2 (\bar{X}_1 - \bar{X}_{01})^2 + S_1^2 (\bar{X}_2 - \bar{X}_{02})^2 - 2S_{12} (\bar{X}_1 - \bar{X}_{01})(\bar{X}_2 - \bar{X}_{02})] \quad (19)$$

$$T_k^2 = (X_k - \hat{\mu})' S^{-1} (X_k - \hat{\mu}) \quad (20)$$

And the S as shows:

$$S = \begin{bmatrix} S_1^2 & S_{12} & \dots & S_{1p} \\ & S_2^2 & \dots & S_{2p} \\ & & \ddots & \vdots \\ & & & S_p^2 \end{bmatrix} \quad (21)$$

Therefore to determine the value of Upper and Lower Limit given By:

$$UCL = \frac{(m-1)^2 [p/(m-p-1)] F_{\alpha, p, m-p-1}}{m + [mp/(m-p-1)] F_{\alpha, p, m-p-1}} \quad (22)$$

$$UCL = \frac{p(m^*+1)(m^*-1)}{(m^*)^2 - m^* p} F_{\alpha, p, m^*-p} \quad (23)$$

m:is the samples, $m=m^*$ if no sample

And the UCL is equal the Value of Chi-Square as:

$$UCL = \chi_{\alpha,p}^2$$

Value of Hotelling's T^2 percentile points can be obtained from the percentile points of the F-distribution given as a relation

$$T_{a,p,n-1}^2 = P\left(\frac{n-1}{n-p}\right)P_{a,p,n-p} \quad (24)$$

Where

$F_{a,p,n-1}$ represents to the F-distribution with p degree of freedom .

EWMA Control charts are used to monitor the mean of a process and are interested in identifying minor shifts. They also incorporate historical data in addition to current data, but the weights attached to the data are exponentially decreasing as the observations become less recent, the new generation of value of Z_i is as:

$$Z_i = \lambda x_i + (1 - \lambda)Z_{i-1} \quad (25)$$

Where $0 < \lambda \leq 1$ is a constant and the starting value required with the first sample at $i=1$ is the process target value, so that

$$Z_0 = u_0$$

Sometimes the average of the data is used to start value, then it become $Z_0 = \bar{x}$ If the observations X_i are independent random variables with σ^2 then the variance is:

$$\sigma_{x_i}^2 = \sigma^2 \sqrt{\frac{\lambda}{2-\lambda} (1 - (1-\lambda)^{2i})} \quad (26)$$

Then the EWMA control chart would be constructed by plotting Z_i versus the sample number i . The centre line and control limits for the EWMA control chart are as follows [6][7][10]:

$$UCL, LCL = \mu_0 \pm L \frac{L\sigma}{\sqrt{n}} \sqrt{\frac{\lambda}{2-\lambda} (1 - (1-\lambda)^{2i})} \quad (27)$$

From the UCL, and LCL that the term $[1 - (1-\lambda)^{2i}]$ Approaches unity as i get larger, than the ULC and LCL it become as:

$$UCL, LCL = \mu_0 \pm L\sigma \sqrt{\frac{\lambda}{2-\lambda}} \quad (28)$$

$$Cl = \mu_0$$

Where σ_x is the standard deviation, where L is the distance between the control limits and the centre line (CL). And λ is the constant such that $0 < \lambda \leq 1$. The quantity Z_0 is the starting value, and it is taken equal to the target mean μ or the average of the initial data in case when the information on the target mean is not available.[3][4][5][11]

4. APPLICATION: NUMERICAL ILLUSTRATION

There are numerical analysis and quality control charts to compare Mahalanobis Distance (MD) and, MEWMA control chart shows the positive measurement aspects of the above presented and demonstrates which one has a good outcome. This paper used data from three chemical components of water to do so. (ACL Temp. (X3), Conductivity (X1), TDS (X2), and Conductivity (X1)).

We have a random sample size (30) for the chemical analysis data for this application using the SPSS and STATGRAPHICS – XVI software and data as in table(1). We determine the value of MD, Multivariate T-Squared Control Chart as shown in table(1).

Table (1) shows Table (1) shows that the value of MD and T-Square.The Max.value of MD is (12.57) and The Max.value for MEWMA is (42.53)

Table 1: contains three kinds of variables of compotes of water.

| Sampl | Conductivity (X1) | TDS (X2) | ACL Temp. C (X3) | EWMA for MD Value | EWMA- For T-Square Value | MD- Value | T-Squared -Value |
|-------|-------------------|----------|------------------|-------------------|--------------------------|-----------|------------------|
| 1 | 183.67 | 117.67 | 15.00 | 4.83 | 15.98 | 12.57 | 42.53 |
| 2 | 187.33 | 120.00 | 17.33 | 4.15 | 14.73 | 1.43 | 9.74 |
| 3 | 189.33 | 121.33 | 18.00 | 3.83 | 13.57 | 2.54 | 8.90 |
| 4 | 191.67 | 122.67 | 17.67 | 3.84 | 13.81 | 3.90 | 14.78 |
| 5 | 189.00 | 120.67 | 15.33 | 4.82 | 18.23 | 8.72 | 35.93 |
| . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . |
| 25 | 168.00 | 107.67 | 19.33 | 2.47 | 8.18 | 3.00 | 11.84 |
| 26 | 169.00 | 108.33 | 19.33 | 2.46 | 8.57 | 2.42 | 10.13 |
| 27 | 174.67 | 112.00 | 20.33 | 2.38 | 9.00 | 2.08 | 10.71 |
| 28 | 173.67 | 111.33 | 20.00 | 2.17 | 8.92 | 1.32 | 8.64 |
| 29 | 170.33 | 109.33 | 19.67 | 2.20 | 9.21 | 2.33 | 10.35 |
| 30 | 177.00 | 113.67 | 19.33 | 2.19 | 8.05 | 2.12 | 3.41 |

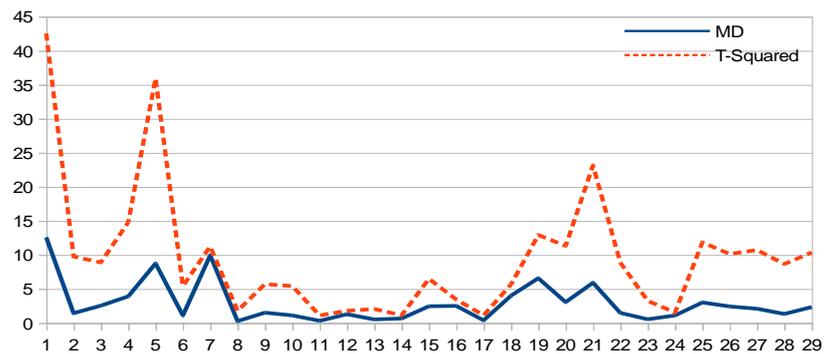


Figure 1: MD and T-Square Chart

Figure (1) shows that the MD and T-Square chart. It is seen that the MD chart has a well-distributed chart, and it is seen that only one point is out of control and the upper control chart of MD is (11.62), as shown in figure (2).

From figure (3) of the Multivariate T-Squared chart, there are (5) points out of control,

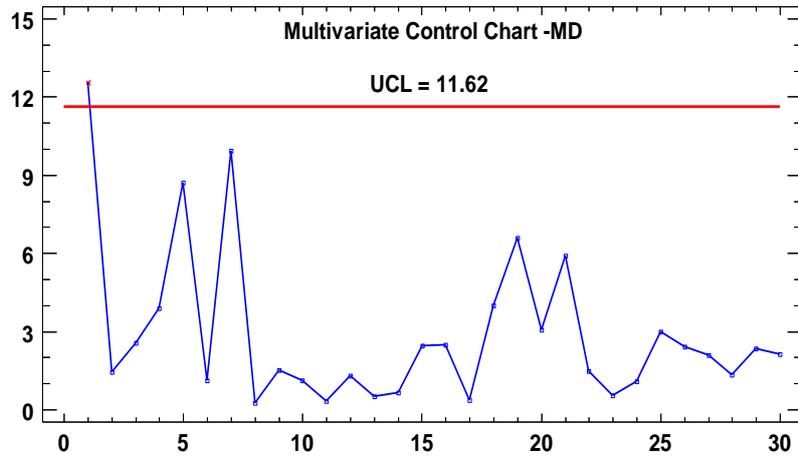


Figure 2: MD Control Chart

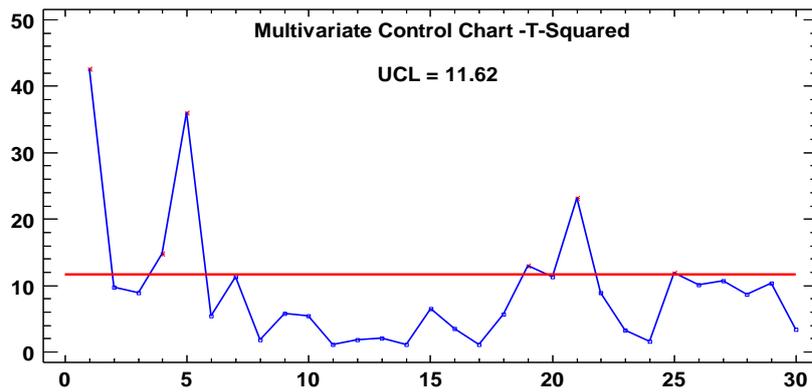


Figure 3: T-Square Chart

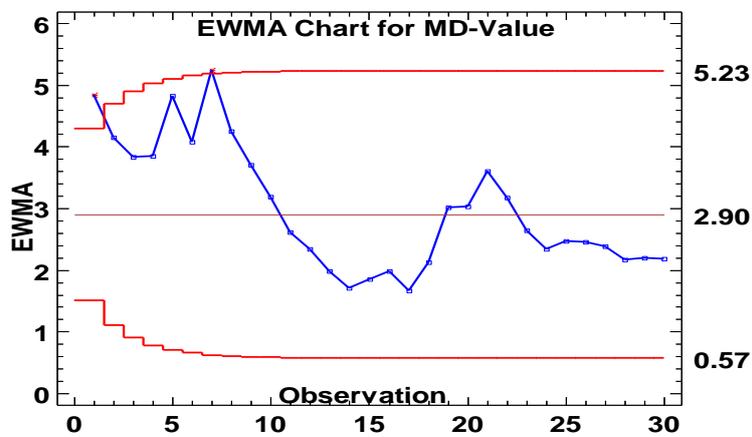


Figure 4:EWMA Chart For MD

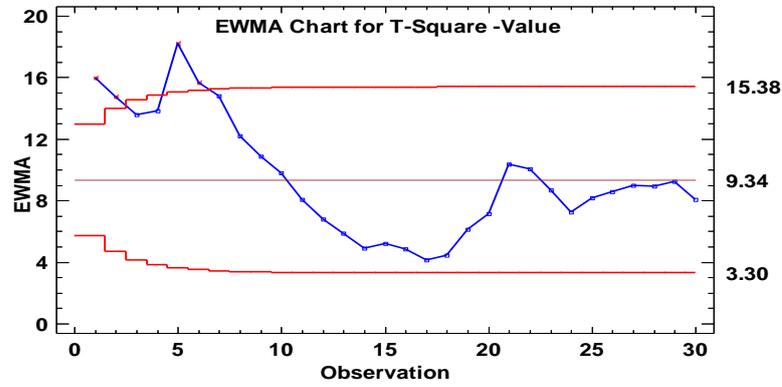


Figure 5: EWMA Chart For T-Square

Figure (4) shows that the EWMA-chart of MD value, it is seen that the UCL, LCL (5.23 ,0.57), and CL is (2.9). There are two points out of control: process capability (2.33), and the UCL, LCL for T-Square (15.38, 3.3). The CL is (9.34), and there are (4) points out of control, with process capability (6.04) as shown in Figure (5).

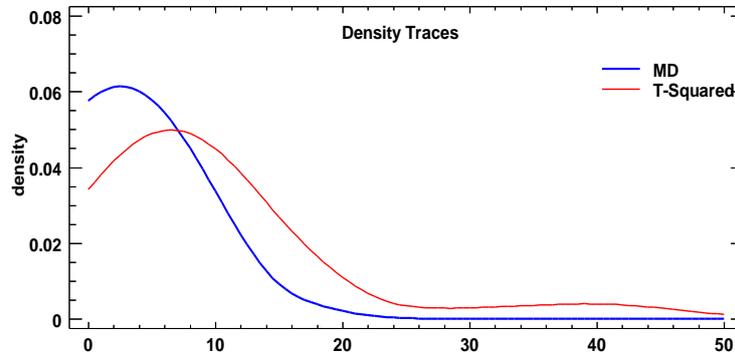


Figure 6: Density of EWMA Chart For T-Square

The density Traces distribution of MD has a better density than T-Square, as shown in Fig(6). The Median of MD is (2.098) and for T-Square is(7.6) by the Whitney-test, there is a statistically significant difference between the medians at the 95.0% confidence level, and the interval for the mean of MD is (2.9 ± 1.12) . The T-square is: (9.34 ± 3.58) with % 95 confidence, And For Variance for MD and T-Square are (9.31,93.84).

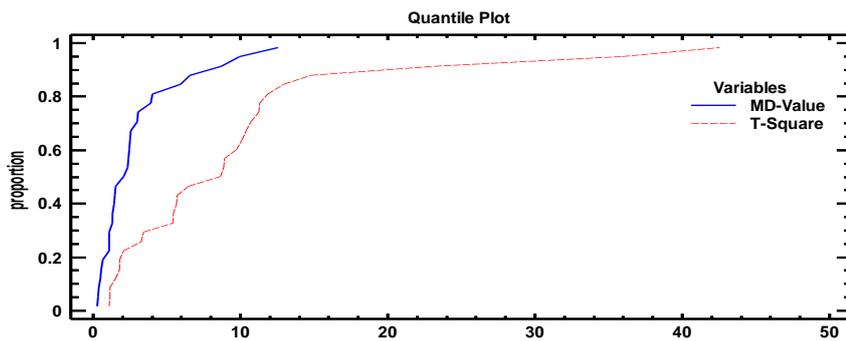


Figure 7: Probability Plot of EWMA Chart For T-Square

The Box whisker Plot has a good separation as shown in Fig(8), the lower and Upper quartile of MD are (1.102,3.26), and for Multivariate T-Squared are (3.05,11.23), with Interquartile range of MD and T-Square are(1.95,7.97) and although for the histogram Frequency. The MD value has a good distribution frequency more than T-Square, as shown in Fig (9).

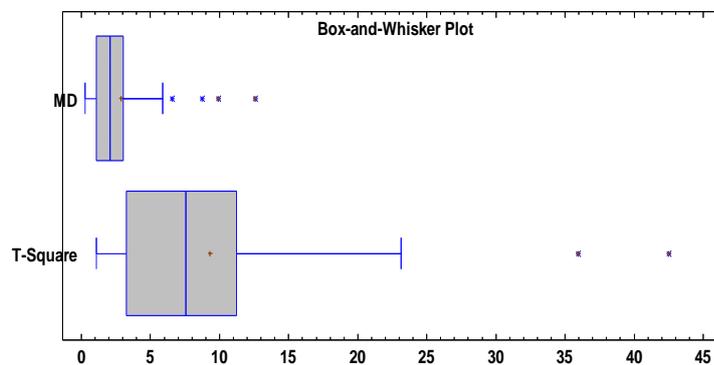


Figure 8: Box Plot of MD and T-Square value

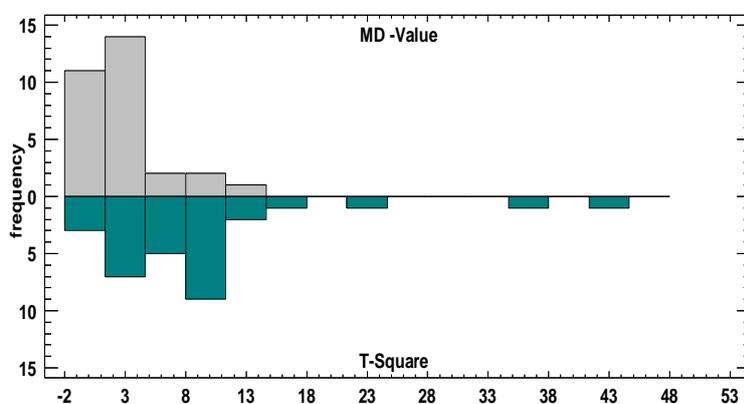


Figure 9: Histogram Frequency of MD and T-Square value

5. CONCLUSION

The main goal of this research paper will be to using the MD-Distance control chart and the Multivariate T-Squared Control Chart. During all Control charts are used. According to the comparison study, a good comparison between the MEWMA and MD-Distance control chart shows that the MD-Distance control chart has a good result. In general, the MD-Distance control chart has a good effect on variables and offers a good relationship between variables.

REFERENCE

- [1] A.Faraz, "Hotelling's T2Control Chart with Double Warning Lines", Quality Management Department, Air Defense Group, Aerospace Industries Organization, Tehran, Iran.
- [2] A. MITRA ,FUNDAMENTALS OF QUALITY CONTROL AND IMPROVEMEN, Fourth Edition Auburn University, College of Business Auburn, Alabama Copyright. 2016 by John Wiley&Sons, Inc. Allrightsreserved.
- [3] A S.T.A. Niaki ,M.Malakib, M.J. Ershadi ,,A particle swarm optimization approach on economic and economic-statistical designs of MEWMA control charts 2011 ,SCIENTIA IRANICA
- [4] D.C. MONTGOMERY 'Introduction to Statistical Quality Control" , seven editions , Arizona State University Printed in the United States of America. 10 9 8 7 6 5 4 3 2 1 ,2013
- [5] K. Yang .J. Trewm,"Multivariate Statistical Methods in Quality Management " Copyright © The McGraw-Hill Companies 2004
- [6] G.TAGUCHI And R. JUGULUM, "The Mahalnobis -Taguchi Strategy" : A Pattern Technology System.Copyright 2002 john Wiley & Sons. Ins.
- [7] K. J.Rashid, 2017," Design of an Exponentially Weighted Moving Average (EWMA) and An Exponential Weighted Root Mean Square (EWRMS) Control Chart, International Journal Of Advanced Engineering Research and Science (IJAERS) [Vol-4, Issue-3, Mar- 2017]
- [8] R. Mawonike, 2017," Multivariate Exponentially Weighted Moving Average for Monitoring Coke Production: A Case of Wankie Colliery Company, Global Journal of Theoretical and Applied Mathematics Sciences. ISSN 2248-9916 Volume 7, Number 1 (2017), pp. 1-16
- [9] S.Bersimis and,J.Panaretos,2005, "Multivariate Statistical Process Control Charts and the Problem of Interpretation: A Short Overview and Some Applications in, University of Economic and Business, Department of Statistics, Athens, Greece
- [10] T.T. Allen., 2010, "Introduction to Engineering Statistics " and Lean Sigma Statistical Quality Control and Design of Experiments and Systems Second Edition.
- [11] T. P. RYAN, Smyrna, Georgia" 2011", Statistical Methods for Quality Improvement, Third Edition Copyright © 2011 by John Wiley & Sons, Inc. All rights reserved. Published by John Wiley & Sons, Inc., Hoboken, New Jersey.