

Decomposing Hotelling's T^2 Statistic Using Four Variables

Nathan S. Agog¹, Hussaini G. Dikko², Osebekwin E. Asiribo³

P.G. Student, Department of Mathematics, Ahmadu Bello University, Zaria, Kaduna State, Nigeria¹

Lecturer, Department of Mathematics, Ahmadu Bello University, Zaria, Kaduna State, Nigeria²

Professor, Department of Mathematics, Ahmadu Bello University, Zaria, Kaduna State, Nigeria³

Abstract: The Mason, Young and Tracy (MYT) decomposition approach has gained much acceptance due to its effectiveness in detecting variables that contributes to an out-of-control condition. The decomposition approach enables one to assess the T^2 value of each variable independently as well as their joint contribution. One of the setbacks of this approach is its computational complexity. In this work, the Hotelling's T^2 decomposition is extended from three variables to four variables using the MYT decomposition technique. Twenty four (24) models were obtained from the decomposition, where each model gives the same T^2 value proving the invariance property of the Hotelling's T^2 statistic.

Keywords: Invariance Property, Matrix Permutation, Multivariate Statistical Process Control (MSPC), MYT decomposition.

I. INTRODUCTION

The decomposition of the Hotelling's T^2 statistic into orthogonal components is considered to be one of the most effective tools for detecting variables that significantly contributes to out-of-control signal. This concept reflects the contribution of each individual variable as well as the joint contribution between two or more variables in a multivariate quality control chart.

Multivariate control charts are widely used in practice to monitor the simultaneous performance of several related quality characteristics. The origin of multivariate control chart can be attributed to Hotelling [1]. A multivariate control scheme has a better sensitivity than the univariate control charts in monitoring multivariate quality process (Lu *et al.*, [2])

Woodall and Montgomery [12] in one of their discussion paper stated that multivariate process control is one of the most rapidly developing sections of statistical process control. The demand to implement MSPC in a production process for quality improvement increases daily. According to Woodall [13], statistical methods play a very important role in quality improvement in manufacturing industries. The most commonly used multivariate statistical technique is the statistical Hotelling's T^2 control chart.

There are a lot of literatures focusing on multivariate control charting methods based on Hotelling's T^2 statistic in detecting mean shift such as Sullivan and Woodall [9], Mason and Young [5], Tong *et al.*, [10] and many others.

Much application of the Hotelling's T^2 decomposition technique has been on two and three process variables as seen in the case of Mason *et al.*, [4], Ulen and Demir [11], Sani and Abubakar [8] and many others.

In this article, our interest is to extend this approach by deriving the T^2 decomposition model using four variables. This decomposition will not only be helpful to quality control engineers but also to researchers in the field of quality control.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2014

II. DECOMPOSITION METHOD

Mason *et al.*, [3], proposed the use of T^2 decomposition in detecting out-of-control variable. The T^2 decomposition is considered to be the most valuable technique in detecting influential variable(s). The main idea of this method is to decompose the T^2 statistic into independent parts, each of which reflects the contribution of an individual variable. The characteristic that is significantly contributing to the signal is more readily identified by decomposing the T^2

Mason *et al.*, [4], presented a practical application of the T^2 decomposition on a bivariate data (two process variables, $p=2$). Where there were two possible decompositions;

$$T^2 = \begin{cases} T_1^2 + T_{2,1}^2 \\ T_2^2 + T_{1,2}^2 \end{cases}$$

Sani and Abubakar [8], presented the decomposition of three process variables ($p=3$) by Mason and Young [6] to demonstrate the invariance property of the Hotelling's T^2 statistic. The decomposition of the Hotelling's T^2 statistic for three process variables is as follows;

$$T^2 = \begin{cases} T_1^2 + T_{2,1}^2 + T_{3,1,2}^2 \\ T_1^2 + T_{3,1}^2 + T_{2,1,3}^2 \\ T_2^2 + T_{3,2}^2 + T_{1,2,3}^2 \\ T_2^2 + T_{1,2}^2 + T_{3,1,2}^2 \\ T_3^2 + T_{1,3}^2 + T_{2,1,3}^2 \\ T_3^2 + T_{2,3}^2 + T_{1,2,3}^2 \end{cases}$$

This work extends the decomposition of Hotelling's T^2 statistic from three variable ($p=3$) to four variables ($p=4$). The number of decompositions increased from 6 to 24 possible ways ($p! = 3! = (6 \text{ decompositions})$ to $p! = 4! = (24 \text{ decompositions})$). These decompositions give the same overall T^2 statistic as stated by Mason *et al.*, [3]

III. MASON, YOUNG AND TRACY (MYT) DECOMPOSITION TECHNIQUE

The Hotelling's T^2 control chart is the most commonly used multivariate control charts when all the quality characteristics are normally distributed. Let x_1, x_2, \dots, x_p be p quality characteristics. Assuming they are normally distributed with multivariate mean, μ and covariance matrix, Σ . The T^2 statistic for a p dimensional observation vector $X' = (x_1, x_2, \dots, x_p)$ can be represented as,

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

This approach is to decompose the T^2 statistic into independent components, each of which reflects the contribution of individual variables. The T^2 statistic in (1) can be partitioned into two independent parts (Rencher [7]). These components are given by,

$$T^2 = T_{p-1}^2 + T_{p,1,2,\dots,p-1}^2 \tag{2}$$

Since the first term in (2) is a T^2 statistic, it can also be separated into two orthogonal parts;

$$T_{p-1}^2 = T_{p-2}^2 + T_{p-1,1,2,\dots,p-1}^2 \tag{3}$$

Continuing in the same fashion, we can generate one of the many possible MYT decompositions of the T^2 statistic.

$$T^2 = T_1^2 + T_{2,1}^2 + T_{3,1,2}^2 + \dots + T_{p,1,2,3,\dots,p-1}^2 \tag{4}$$

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2014

III.1. T² DECOMPOSITION MODEL USING FOUR VARIABLES

The process is initiated by selecting any one of the p variables, followed by any of the $(p-1)$ remaining variables to condition on the first selected variable. Next, select any of the remaining $(p-1)$ variables to be condition on the first two selected variables. And finally, select any of the $(p-3)$ variables to condition on the first three selected variables. Iterating in the same procedure will generate all the decomposition equations which compose the same over all T² statistic.

For the first p variables, assuming we select $p=1$ (remaining 2nd, 3rd and 4th variables). The decomposition is as follows;

$$T^2 = T_1^2 + T_{2,1}^2 + T_{3,1,2}^2 + T_{4,1,2,3}^2$$

$$T^2 = T_1^2 + T_{3,1}^2 + T_{4,1,3}^2 + T_{2,1,3,4}^2$$

$$T^2 = T_1^2 + T_{4,1}^2 + T_{2,1,4}^2 + T_{3,1,2,4}^2$$

The next step is to keep the selected p variable and the $(p-1)$ remaining variables conditioned on the p selected variable constant (that is, the first and the second decomposition terms). Then select any of the remaining $(p-2)$ variables not used in the first decomposition to be condition on the first two kept variables. And finally, select any of the remaining $(p-3)$ variables not used during the first decomposition to condition on the first three selected variables. These yield the following results;

$$T^2 = T_1^2 + T_{2,1}^2 + T_{4,1,2}^2 + T_{3,1,2,4}^2$$

$$T^2 = T_1^2 + T_{3,1}^2 + T_{2,1,3}^2 + T_{4,1,2,3}^2$$

$$T^2 = T_1^2 + T_{4,1}^2 + T_{3,1,4}^2 + T_{2,1,3,4}^2$$

Iterating in the same procedure will generate all the possible decomposition equations which compose the same overall T² statistic. This condition satisfies the invariance property which states that the value of the T² cannot change by permuting the observation vector (Mason and Young [6]). Hence, the complete T² decomposition using four variables ($p = 4$) is as follows;

$$T^2 = T_1^2 + T_{2,1}^2 + T_{3,1,2}^2 + T_{4,1,2,3}^2$$

$$T^2 = T_1^2 + T_{3,1}^2 + T_{4,1,3}^2 + T_{2,1,3,4}^2$$

$$T^2 = T_1^2 + T_{4,1}^2 + T_{2,1,4}^2 + T_{3,1,2,4}^2$$

$$T^2 = T_1^2 + T_{2,1}^2 + T_{4,1,2}^2 + T_{3,1,2,4}^2$$

$$T^2 = T_1^2 + T_{3,1}^2 + T_{2,1,3}^2 + T_{4,1,2,3}^2$$

$$T^2 = T_1^2 + T_{4,1}^2 + T_{3,1,4}^2 + T_{2,1,3,4}^2$$

$$T^2 = T_2^2 + T_{1,2}^2 + T_{3,1,2}^2 + T_{4,1,2,3}^2$$

$$T^2 = T_2^2 + T_{3,2}^2 + T_{4,2,3}^2 + T_{1,2,3,4}^2$$

$$T^2 = T_2^2 + T_{4,2}^2 + T_{1,2,3}^2 + T_{3,1,2,4}^2$$

$$T^2 = T_2^2 + T_{1,2}^2 + T_{4,1,2}^2 + T_{3,1,2,4}^2$$

$$T^2 = T_2^2 + T_{3,2}^2 + T_{1,2,3}^2 + T_{4,1,2,3}^2$$

$$T^2 = T_2^2 + T_{4,2}^2 + T_{3,2,4}^2 + T_{1,2,3,4}^2$$

$$T^2 = T_3^2 + T_{1,3}^2 + T_{2,1,3}^2 + T_{4,1,2,3}^2$$

$$T^2 = T_3^2 + T_{2,3}^2 + T_{4,2,3}^2 + T_{1,2,3,4}^2$$

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2014

$$\begin{aligned}
 T^2 &= T_3^2 + T_{4.3}^2 + T_{1.3,4}^2 + T_{2.1,3,4}^2 \\
 T^2 &= T_3^2 + T_{1.3}^2 + T_{4.1,3}^2 + T_{2.1,3,4}^2 \\
 T^2 &= T_3^2 + T_{2.3}^2 + T_{1.2,3}^2 + T_{4.1,2,3}^2 \\
 T^2 &= T_3^2 + T_{4.3}^2 + T_{2.3,4}^2 + T_{1.2,3,4}^2 \\
 T^2 &= T_4^2 + T_{1.4}^2 + T_{2.1,4}^2 + T_{3.1,2,4}^2 \\
 T^2 &= T_4^2 + T_{2.4}^2 + T_{3.2,4}^2 + T_{1.2,3,4}^2 \\
 T^2 &= T_4^2 + T_{3.4}^2 + T_{1.3,4}^2 + T_{2.1,3,4}^2 \\
 T^2 &= T_4^2 + T_{1.4}^2 + T_{3.1,4}^2 + T_{2.1,3,4}^2 \\
 T^2 &= T_4^2 + T_{2.4}^2 + T_{1.2,4}^2 + T_{3.1,2,4}^2 \\
 T^2 &= T_4^2 + T_{3.4}^2 + T_{2.3,4}^2 + T_{1.2,3,4}^2
 \end{aligned}$$

IV. COMPUTING THE T² DECOMPOSITION TERMS

One of the decomposition for the observation vector (x_1, x_2, x_3, x_4) is given as

$$T_{(x_1, x_2, x_3, x_4)}^2 = T_1^2 + T_{3.1}^2 + T_{2.1,3}^2 + T_{4.1,2,3}^2 \tag{5}$$

The computation starts by first determining the value of the conditional term, $T_{4.1,2,3}^2$. From the above equation,

$$T_{4.1,2,3}^2 = T_{(x_1, x_2, x_3, x_4)}^2 - T_{(x_1, x_2, x_3)}^2 \tag{6}$$

The variance-covariance matrix and the mean vector for the observation vector (x_1, x_2, x_3, x_4) is given as;

$$S_{44} = \begin{pmatrix} S_1^2 & S_{12} & S_{13} & S_{14} \\ S_{21} & S_2^2 & S_{23} & S_{24} \\ S_{31} & S_{32} & S_3^2 & S_{34} \\ S_{41} & S_{42} & S_{43} & S_4^2 \end{pmatrix} \text{ and } \bar{X}^{(4)} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \\ x_4 \end{pmatrix} \tag{7}$$

Thus the computation of $T_{(x_1, x_2, x_3, x_4)}^2$ is as follows;

$$T_{(x_1, x_2, x_3, x_4)}^2 = (X^{(4)} - \bar{X}^{(4)})' S_{44}^{-1} (X^{(4)} - \bar{X}^{(4)}) \tag{8}$$

To obtain $T_{(x_1, x_2, x_3)}^2$, the original estimates of the mean vector and covariance structure is partitioned to obtain the mean vector and covariance matrix of the sub vector $X^{(3)} = (x_1, x_2, x_3)$. The corresponding partition is given as;

$$S_{33} = \begin{pmatrix} S_1^2 & S_{12} & S_{13} \\ S_{21} & S_2^2 & S_{23} \\ S_{31} & S_{32} & S_3^2 \end{pmatrix} \text{ and } \bar{X} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ x_3 \end{pmatrix} \tag{9}$$

Thus the computation $T_{(x_1, x_2, x_3)}^2$ is as follows;

$$T_{(x_1, x_2, x_3)}^2 = (X^{(3)} - \bar{X}^{(3)})' S_{33}^{-1} (X^{(3)} - \bar{X}^{(3)}) \tag{10}$$

Also the decomposition of $T_{(x_1, x_2, x_3)}^2$ is given by

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2014

$$T_{(x_1, x_2, x_3)}^2 = T_1^2 + T_{3.1}^2 + T_{2.1,3}^2 \tag{11}$$

The equation above can be obtained by first computing the conditional term $T_{2.1,3}^2$ as follows;

$$T_{2.1,3}^2 = T_{(x_1, x_2, x_3)}^2 - T_{(x_1, x_3)}^2 \tag{12}$$

To obtain the term $T_{(x_1, x_3)}^2$, the original estimates of the mean vector and the covariance structure is partitioned to obtain the mean vector and covariance matrix of the sub vector $X^{(2)} = (x_1, x_3)$. The corresponding partition is given as;

$$S_{22} = \begin{pmatrix} S_{11}^2 & S_{13} \\ S_{31} & S_3^2 \end{pmatrix} \text{ and } \bar{X}^{(2)} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_3 \end{pmatrix} \tag{13}$$

Hence, the computation of the term $T_{(x_1, x_3)}^2$ is as follows;

$$T_{(x_1, x_3)}^2 = (X^{(2)} - \bar{X}^{(2)})' S_{22}^{-1} (X^{(2)} - \bar{X}^{(2)}) \tag{14}$$

Also, the decomposition for $T_{(x_1, x_3)}^2$ is given by

$$T_{(x_1, x_3)}^2 = T_1^2 + T_{3.1}^2 \tag{15}$$

The term $T_{3.1}^2$ is obtained by computing the T^2 value of the sub vector $X^{(1)} = (x_1)$. Hence, the unconditional term T_1^2 is computed by

$$T_1^2 = \frac{(x_1 - \bar{x}_1)^2}{s_1^2} \tag{16}$$

Thus, $T_{3.1}^2$ is computed as;

$$T_{3.1}^2 = T_{(x_1, x_3)}^2 - T_{(x_1)}^2 \tag{17}$$

$$T_j^2 = \frac{(x_j - \bar{x}_j)^2}{s_j^2} \tag{18}$$

$j=1,2,\dots,p$. is the square of a univariate t statistic for the observed value of the j^{th} variable of vector X .

V. CONCLUSION

The MYT decomposition has a beautiful way of presenting the contribution of independent variables and also the relationship between two or more variables. This work has presented a simplified way decomposing the T^2 statistic by using four process variables where 24 decompositions were obtained. Each of these decompositions can be used to identify variable(s) that significantly contributes to an out-of-control signal.

REFERENCES

1. Hotelling, H. "Multivariate quality control. In C. Eisenhart, M.W. Hastay and W.A. Wallis, eds.", Techniques of Statistical Analysis. New York: McGraw Hill, pp 111-184,1947.
2. Lu, X.S., M. Xie, T.N. Goh and C.D. Lai, "Control chart for multivariate attribute Processes", International Journal Production Research, Vol. 36, pp. 3477-3489, 1998.
3. Mason R.L., Tracy N.D. and Young J.C., "Decomposition of T^2 for multivariate control chart interpretation" Journal of Quality Technology Vol. 27, pp. 99-108,1995.
4. Mason, R. L., Tracy N.D., and Young J.C. A practical application for interpreting multivariate T^2 control chart signal . Journal of Quality Technology, Vol. 29, No.4, pp.396-501, 1997.
5. Mason, R.L. and Young, J.C.. Improving the sensitivity of the T^2 statistic in multivariate process control. Journal of Quality Technology, Vol. 31, No. 2, pp. 155-165, 1999.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2014

6. Mason, R.L. and Young, J.C., "Multivariate Statistical Process Control with Industrial Applications", Philadelphia: ASA-SIAM. pp. 146-161, 2002
7. Rencher, A.C., "The contribution of individual variables to Hotelling's T^2 , Wilks' Λ and R^2 ", Biometrics, Vol. 49, pp. 479-489, 1993.
8. Sani, S.A., and Abubakar, Y., "Demonstrating the Invariance Property of the Hotelling T^2 Statistic", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 2, No. 7, pp. 2515-2519, 2013.
9. Sullivan, J.H. and Woodall, W.H., "A comparison of multivariate control charts for individual observations", Journal of Quality Technology, Vol. 28, No. 4, pp. 398-408, 1996.
10. Tong, L.I., Wang, C.H. and Huang, C.L., "Monitoring defects in IC fabrication using a Hotelling T^2 control chart", IEE Trans. Semiconductor Manufacturing, Vol. 18, No. 1, pp.140-148, 2005.
11. Ulen, M. and Demir, I., "Application of multivariate statistical quality control in pharmaceutical industry", Balkan Journal of Mathematics, Vol. 1, pp. 93-105, 2013.
12. Woodall, W.H and Montgomery, D.C., Research issues and ideas in statistical process control. Journal of Quality Technology, Vol. 31, No. 4, pp. 376-386, 1999.
13. Woodall, W.H., "Controversies and contradictions in statistical process control", Journal of Quality Technology, vol. 32, No.4, pp. 341-350, 2000.