

DETECTING ASSIGNABLE SIGNALS via DECOMPOSITION OF T^2 STATISTIC

Sani Salihu Abubakar¹, Osebekwin Ebenezer Asiribo², Abubakar Yahaya³, Hussaini Garba Dikko⁴

Lecturer, Department of Statistics, Kano university of Science and Technology, Wudil, Nigeria¹

Professor, Department of Statistics, University of Agriculture, Abeokuta, Nigeria²

Dr. Department of Mathematics, Ahmadu Bello University Zaria, Nigeria³

Dr. Department of Mathematics, Ahmadu Bello University Zaria, Nigeria⁴

Abstract: The proliferations of the various means through which data can be accessed technologically via computers have led to a huge enthusiasm in the simultaneous monitoring of several correlated quality characteristics. Hotelling's T^2 is a famous statistic that is so significant in identifying a shift that is due to the mean vector of a multivariate normal distribution. When there is signals in a multivariate control chart, it is a clear indication that special cause variation exist in the process, as such effort must be made to identified which variable (s) is/are responsible for the out of control condition. In this study, we have demonstrated the use of mason, Young and Tracy (MYT) approach in identifying non-random quality characteristics.

Keywords: *Quality control, Multivariate process control, control limits, Hotelling's T^2 statistic.*

I. INTRODUCTION

In 1920's Walter Shewart, was first to developed statistical control chart, essentially, in quality control the importance of control chart can hardly be over emphasized. This fact could be traced to the history of World War II, control charts were readily used during the war and has been employed with various modifications ever since. Jackson [4].

Multivariate statistical process control (SPC) using Hotelling's T^2 statistic is commonly employed to detect shift. However, T^2 control chart has a limitation as it cannot point to the causes of the change. Thus, decomposition of T^2 is recommended, aimed at providing a way of identifying the variables significantly contributing to an out of control signals.

1.1. Shewhart Multivariate Control Charts

Broadly speaking, in the literature, two different phases of control chart have been extensively discussed. Bersimis et al [1].

Phase I: In this phase charts are used for retrospective test to see whether the process was in control when the first subgroups were drawn. The charts are regarded as tools which would aid the practitioner in ensuring that the process is statistically in control. Having accomplished this, the control chart can be employed to define what is meant by a process being statistically in control. According to Duncan [2], phase I ensure the establishment of a process being statistically in control.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

Phase II: In this phase, the charts are to be used in testing whether the process remains in control when future subgroups are drawn. Similarly, the charts here serves as an aid to quality control practitioner in monitoring the process for any change from an in control state, irrespective of whether the parameters of the process μ_0 and Σ_0 , were known or estimated.

Woodall [13], stated that, the transition from phase I to phase II pertaining process understanding and process improvement required a lot of work. Sparks [11], Wierda [12], Lowry and Montgomery [5], Fuchs and Kenett [3], Ryan [9], and other statisticians and engineers share the same view.

1.2. Sample Size

With regards to the issue of sample size n , of each rational subgroup, four approaches have been suggested, with the use of either λ^2 or T^2 test statistic by Lowry and Montgomery [5], as follows:

- (1) Phase I and $n = 1$, working with individual observations;
- (2) Phase I and $n > 1$, working with rational subgroups;
- (3) Phase II and $n = 1$, working with individual observations;
- (4) Phase II and $n > 1$, working with rational subgroups.

It is also imperative to note that, Mason and Young [7], highlighted the basic steps towards implementation of multivariate statistical process control using the T^2 statistic.

In the subsequent sections, the article would deal on control chart for individual observation, thereafter; we shall present the decomposition techniques as proposed by Mason, Tracy and Young (MTY), also in order to enhance the reader's understanding of the technique an illustrative example would be given. Finally, results discussion and conclusion would emerge.

II. CONTROL CHART FOR INDIVIDUAL OBSERVATION

Usually, a statistic that is most frequently employed in multivariate control charts for individual observation is Hotelling's T^2 (Mason and Young,) [6], which is defined as the generalized distance from a p -dimensional sample point $X = (x_1, x_2, \dots, x_p)'$ to its sample mean. It is given by:

$$T^2 = (X - \bar{X})' S^{-1} (X - \bar{X}) \text{-----} (2.1)$$

Where \bar{X} and S are the common estimators for the mean vector and covariance matrix obtained from historical data set. When the observation vector, X is independent of the estimates \bar{X} and S , the distribution of the statistic in (2.1) follows an F-distribution i.e.

$$\frac{n(n-p)T^2}{p(n+1)(n-1)} \square F(\alpha, p, n-p) \text{-----} (2.2)$$

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

2.1 The MYT Decomposition

The T^2 statistic in (2.1) can be split or decomposed into p-orthogonal components (Mason, Tracy and Young,) [8], for instance, if you have p- variates to decompose, the procedure is as follows:

$$T^2 = T_1^2 + T_{2,1}^2 + \dots + T_{p,1,2,\dots,p-1}^2 \quad (2.3)$$

The first term is an unconditional Hotelling's T^2 , decomposing it as the first variable of the observation vector X_1 as:

$$T_1^2 = \frac{(x_1 - \bar{x}_1)^2}{s_1^2} \quad (2.4)$$

where \bar{x} and s_1 is the mean and standard deviation of the variable X_1 respectively. T_j^2 will follow an F-distribution which can be used as upper control limit.

$$T_j^2 = \frac{(x_j - \bar{x}_j)^2}{s_j^2} \quad \square \quad \frac{n+1}{n} F_{\alpha, 1, n-1},$$

Taking a case of three variables as an example, it can be decomposed as,

$$T^2 = \begin{cases} T_1^2 + T_{2,1}^2 + T_{3,12}^2 \\ T_1^2 + T_{3,1}^2 + T_{2,13}^2 \\ T_2^2 + T_{1,2}^2 + T_{3,12}^2 \\ T_2^2 + T_{3,2}^2 + T_{1,23}^2 \\ T_3^2 + T_{1,3}^2 + T_{2,13}^2 \\ T_3^2 + T_{2,3}^2 + T_{1,23}^2 \end{cases} \quad (2.5)$$

It is obvious that with increase in the number of variables, the number of terms will also increase drastically which makes the computation become tedious.

III. ILLUSTRATIVE EXAMPLE

In this section, we intend to demonstrate by way of example, how an assignable signal could be detected, the data that was used for the purpose of the research was collected from a West African Portland cement company in Lagos, Nigeria. The data comprises readings of temperature from a boiler machine in which twenty five observations were taken under three different temperatures. The data can be found in appendix [1], as in Sani and Abubakar [10] the covariance matrix and the mean vector below were obtained from the data.

$$S = \begin{pmatrix} 54 & 0.958 & 20.583 \\ 0.958 & 4.84 & 2.963 \\ 20.583 & 2.963 & 22.993 \end{pmatrix}, \quad \bar{X} = \begin{pmatrix} 525 \\ 513.36 \\ 538.92 \end{pmatrix}$$

Having obtained the covariance matrix and the mean vector, the computation of T^2 statistic was done using mat lab computer package and the values obtained for each of the twenty five observations are as shown in the table below.

**International Journal of Innovative Research in Science,
Engineering and Technology**

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

TABLE 1:

Computation of Hotelling's T^2 for Individual Observation

No. of observation	Hotelling's T^2 value	Critical value F control function
1	10.6950**	10.38
2	3.1972	
3	0.9558	
4	0.5567	
5	0.8560	
6	1.3567	
7	0.2384	
8	4.4559	
9	15.7660**	
10	1.2997	
11	1.1087	
12	0.3429	
13	0.4584	
14	5.3476	
15	1.7978	
16	0.8790	
17	1.9216	
18	2.0612	
19	4.6490	
20	3.2212	
21	0.4959	
22	1.0516	
23	4.1616	
24	0.8069	
25	4.3591	

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

Below is the graph of Hotelling's T^2 for individual observation, where the individual T^2 values were plotted against the number of observations. And the dotted line represents the control limit.

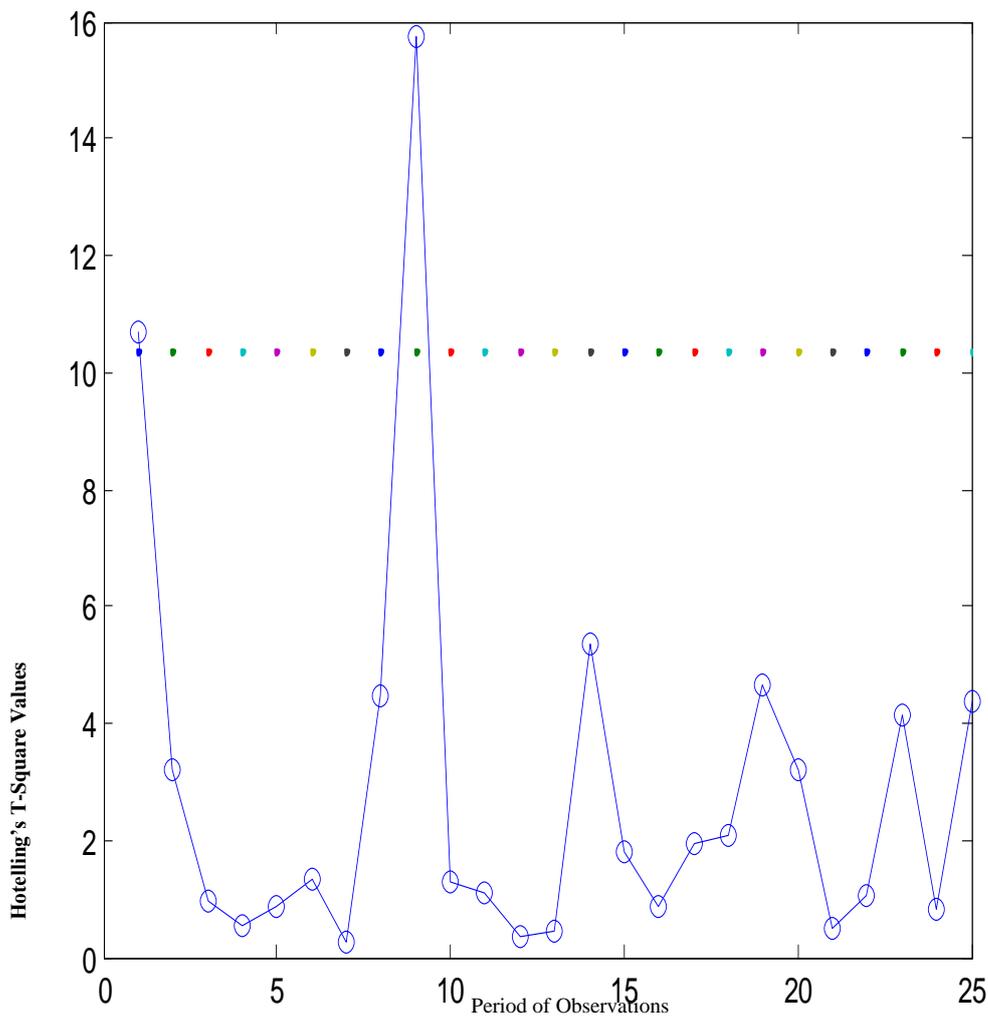


Figure 1: Chart of Hotelling's T-Square Statistic

As it can be seen from the graph as well as the computation of T^2 statistic in the table above, it is vivid that observations 1 and 9 were out of control, as such, in order to detect the assignable signal, we consider observation 9, and the procedure is as follows:

Given that $X' = (533, 514, 528)$ and $\bar{X} = \begin{pmatrix} 525 \\ 513.36 \\ 538.92 \end{pmatrix}$

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

**International Journal of Innovative Research in Science,
Engineering and Technology**

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

$$T^2(x_1, x_2, x_3) = (8, 0.44, -10.92) \begin{pmatrix} 0.029 & 0.0109 & -0.0270 \\ 0.0109 & 0.228 & -0.039 \\ -0.0270 & -0.039 & 0.073 \end{pmatrix} \begin{pmatrix} 8 \\ 0.44 \\ -10.92 \end{pmatrix} = 15.77$$

$$UCL = \frac{p(n+1)(n-1)}{n(n-p)} F_{\alpha, p, n-p} = 10.38$$

From the result above the T^2 overall is significant, as such we now decompose the T^2 into its orthogonal components as to be able to find out which among the three variables contribute most as well as responsible to the out of control signals.

Recall from equation (2.5) above, where for $p = 3$ we had 18 components of T^2 with 6 different equations each yielding the same overall T^2 .

The calculation for each component in (2.5) as well as the sum of each term is as follows:

Starting with unconditional terms we have:

$$T_1^2 = \frac{(x_1 - \bar{x}_1)^2}{s_{11}} = 1.19, \text{ Also, } T_2^2 = \frac{(x_2 - \bar{x}_2)^2}{s_{22}} = 0.04 \text{ and}$$

$$T_3^2 = \frac{(x_3 - \bar{x}_3)^2}{s_{33}} = 5.19$$

Now the UCL for the unconditional terms is computed as follows:

$$UCL(x_j) = \frac{n+1}{n} F_{\alpha, 1, n-1} = 4.4304$$

It is evident that T_1^2 and T_2^2 are in control, however, $T_3^2 = 5.19 > 4.4304$, implying that the variable x_3 is part of signal contained in the vector of observation.

To compute the conditional terms we proceed as follows:

$$T_{3,1,2}^2 = T^2(x_1, x_2, x_3) - T^2(x_1, x_2)$$

To obtain $T^2(x_1, x_2)$, separate the original estimates of the mean vector and covariance structure in order to obtain the mean vector and covariance matrix of the sub vector as:

$X^{(2)} = (x_1, x_2)$. The results of the separations are given as:

$$S_{22} = \begin{pmatrix} 54 & 0.958 \\ 0.958 & 4.84 \end{pmatrix}, \bar{X}^{(2)} = \begin{pmatrix} 525 \\ 513.36 \end{pmatrix}. \text{ We compute}$$

$$T^2(x_1, x_2) = (x^{(2)} - \bar{x}^{(2)})' s_{22}^{-1} (x^{(2)} - \bar{x}^{(2)})$$

$$T^2(x_1, x_2) = 1.21. \text{ Thus, } T_{2,1}^2 = T^2(x_1, x_2) - T_1^2 = 0.02 \text{ And}$$

$$T_{3,1,2}^2 = T^2(x_1, x_2, x_3) - T^2(x_1, x_2) = 14.56$$

From the above, we have:

$$T^2(x_1, x_2, x_3) = T_1^2 + T_{2,1}^2 + T_{3,1,2}^2 = 15.77$$

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

For the fact that the first two terms have smaller values implies that the signal is contained in the third term $T_{3,1,2}^2$. Next, to check whether there is signal in (x_1, x_2) , we partition the original observations into two groups (x_1, x_2) and x_3 .

Having computed the value of $T^2(x_1, x_2) = 1.21$ we compare it to the

$$UCL(x_1, x_2) = 7.4229$$

We conclude that, there is no signal present in (x_1, x_2) component of the observation vector. Hence the reason for the signal lies in the third component, $x_3 = 528$.

The above decomposition considers just one possible MYT, so as to illustrate the computing technique for the various decompositions terms. It's guaranteed that, whichever MYT decomposition is taken, it will surely yield the same overall T^2 as follows:

$$T_1^2 + T_{2,1}^2 + T_{3,1,2}^2 = 1.19 + 0.02 + 14.56 = 15.77$$

$$T_1^2 + T_{3,1}^2 + T_{2,1,3}^2 = 1.19 + 12.88 + 1.7 = 15.77$$

$$T_2^2 + T_{1,2}^2 + T_{3,1,2}^2 = 0.04 + 1.17 + 14.56 = 15.77$$

$$T_2^2 + T_{3,2}^2 + T_{1,2,3}^2 = 0.04 + 5.91 + 9.82 = 15.77$$

$$T_3^2 + T_{1,3}^2 + T_{2,1,3}^2 = 5.19 + 8.88 + 1.7 = 15.77$$

$$T_3^2 + T_{2,3}^2 + T_{1,2,3}^2 = 5.19 + 0.76 + 9.82 = 15.77$$

With the results of the decompositions above, the value of each term of the decomposition was compared to the respective critical values as depicted in the table below:

TABLE 2:

MYT Decomposition Components for Observation (9)

Components	Value	Critical Value
T_1^2	1.19	4.4304
T_2^2	0.04	4.4304
T_3^2	5.19**	4.4304
$T_{1,2}^2$	1.17	7.4229
$T_{1,3}^2$	8.88**	7.4229
$T_{2,1}^2$	0.02	7.4229
$T_{3,1}^2$	12.88**	7.4229
$T_{2,3}^2$	0.76	7.4229
$T_{3,2}^2$	5.91	7.4229
$T_{1,2,3}^2$	9.82	10.38
$T_{2,1,3}^2$	1.7	10.38

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

$T_{3.1,2}^2$	14.56**	10.38
---------------	---------	-------

From the table above, only one of three unconditional T^2 terms associated with observation (9) has significant value.

Also, looking at the bivariate statistic, we discovered that $T_{1,3}^2$ and $T_{3,1}^2$ have significant values, which means there is a problem in the relation between x_1 and x_3 . So we may conclude that both x_1 and x_3 are potential causes of the abnormal observation.

To rectify the problem, we remove x_1 and x_3 from the vector of observation (9) and check whether the sub vector signals or not.

$$T^2 - T_1^2 - T_3^2 = 15.77 - 1.19 - 5.19 = 9.39 < 10.38$$

The outcome portrays that, the sub vector is insignificant.

The meaning of significant value of $T_{1,3}^2$ is that x_1 conditioned by x_3 deviates from the variable relation pattern established from the historical data. Similarly, and vice-versa for the meaning of significant value of $T_{3,1}^2$, finally x_1 and x_3 were detected as being responsible sources for the assignable signals in observation (9).

IV. CONCLUSION

In this article, we were able to show how an assignable signal could be detected with the aid of Hotelling's T^2 statistic decomposition procedure. Basically control charts are employed so as to be able to distinguish between natural and assignable causes in the variability of quality characteristics. In view of this, identifying which combination of quality characteristic is responsible for the signal of the control chart will ultimately help us know the appropriate actions needed to be taken in order to improve the quality.

REFERENCES

- [1] Bersimis, S., Psarakis, S., and Panaretos, J. (2007) Multivariate Statistical Process Control Charts: an overview, *Quality and Reliability Engineering international*.
- [2] Duncan A.J (1986) Quality control and industrial statistics 5th edn. Irwin: Homewood, IL.
- [3] Fuchs C, Kenett R.S. (1998) Multivariate Quality control. Marcel Dekker, New York.
- [4] Jackson, J. E. (1985). Multivariate Quality Control. *Communication in Statistics-Theory and Methods*. 14:2657-2688
- [5] Lowry, C.A. and Montgomery, D. C. (1995). A review of Multivariate Control Chart. *IIE Transactions*. 27:800-810
- [6] Mason R.L. and Young, J.C. (2002) Multivariate Statistical Process Control with Industrial Applications, ASA-SIAM, Philadelphia.
- [7] Mason R.L. Young J.C. (2001) Implementing Multivariate Process Control Using Hotelling's T^2 Statistic. *Quality progress* 34(4):71-73
- [8] Mason, R.L., Tracy, N.D., and Young, J. C. (1995). Decomposition of T^2 for Multivariate Control Chart Interpretation. *Journal of Quality Technology*. 27: 99-108.
- [9] Ryan T.P. (2000) Statistical Methods for Quality Improvement. Wiley: New York.
- [10] Sani Salihu Abubakar and Abubakar Yahaya (2013). Demonstrating the invariance property of Hotelling's T^2 Statistic. *International Journal of Innovative Research in Science, Engineering and Technology*. Vol. 2, Issue 7
- [11] Sparks R.S (1992) Quality control with multivariate data. *Australian journal of statistics* 34:375-390.
- [12] Wierda S.J. (1994) Multivariate statistical process control- recent results and directions for future research. *Statistica Neerlandica* 48: 147-168
- [13] Woodall W.H. (2000) Controversies and contradictions in Statistical Process Control. *Journal of Quality Technology* 32:341-350

**International Journal of Innovative Research in Science,
Engineering and Technology**

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

APPENDIX 1
Boiler Temperature Data

Periods of Observations	X_1	X_2	X_3
1	507	516	527
2	512	513	533
3	520	512	537
4	520	514	538
5	530	515	542
6	528	516	541
7	522	513	537
8	527	509	537
9	533	514	528
10	530	512	538
11	530	512	541
12	527	513	541
13	529	514	542
14	522	509	539
15	532	515	545
16	531	514	543
17	535	514	542
18	516	515	537
19	514	510	532
20	536	512	540
21	522	514	540
22	520	514	540
23	526	517	546
24	527	514	543
25	529	518	544
Total	13125	12839	13473
Average	525	513.36	538.92