International Journal of Latest Research in Engineering and Management" (IJLREM)
ISSN: 2456-0766
www.ijlrem.org Volume 1 Issue 1 ǁ November. 2017 ǁ PP 01-07

Detecting an Out-Of-Control Signal in A Multivariate Quality Control Chart for Individual Observations

[1,]Nura Muhammad , [2,]Ambi Polycarp Nagwai,
[1,,]Lecturer, Department of Computer and Statistics, Kano State polytechnic, Kano, Nigeria[1]
[2,]Professor, Department of Statistics, University of Ewe, Ghana[2]

ABSTRACT: *The Hotelling'sT^{2}Chart is used in Process monitoring when a quick detection of shifts in the mean vector is desired. When there are shifts in a multivariate Control Charts, it is clearly shows that special cause variation is present in the Process, but the major drawback of Hotelling's T^2 is inability to identify which variable(s) is/are the source of the signals. Hence effort must be made to identify which variable(s) is/are responsible for the out- of-Control situation. In this article, we employ Mason, Young and Tracy (MYT) approach in identifying the variables for the signals.*

KEYWORDS: *Quality Control, Multivariate Statistical Process Control, Hotelling's T^2 Statistic.*

I. INTRODUCTION:

Nowadays, One of the most powerful tools in Quality Control is the Statistical Control Chart developed in the 1920s by Walter She wharts, the Control Chart found wide spread use during World War II and has been employed, with various modifications, ever since. Multivariate Statistical Process Control (SPC) using Hotelling's T^2 statistic is usually employed to detect shifts. However, Hotelling's T^2ControlChart has a shortcoming as it can't figure out the causes of the change. Thus, decomposition ofT^2 is recommended and aims at paving a way of identifying the variable(s) significantly contributing to an out- of-Control signals.

Multivariate Statistical Process Control: The Shewhart control charts have been widely applied in a variety of industries because it is very Simple to implement and the information generated from the Shewhart control charts is also easyfor plant staff to understand. However, monitoring each process variable with separate Shewhart Control chart ignores the correlation between variables and does not fully reflect the real process situation. Nowadays, the process industry has become more complex than it was in the past andinevitably that number of process variables need to be monitored has increased dramatically. It's very often, these variables are multivariate in nature and using Shewhart control charts becomes insufficient.

Hotelling's T^2 Statistic : Hotelling H. (2) can be viewed as the originator of multivariate control charts. Hotelling Proposed a concept of generalized distance between new observations to its sample mean. We first illustrate how this method works with a Multivariate case. Assuming these X_1, X_2 *and* X_3 are distributed according to the Multivariate normal distribution. Let represent a *p* dimensional vector of measurements made on a process. Assuming that when the process is in control, the $X_i$are independent and follow a multivariate normal distribution with mean vector μ and covariance matrix Σ. Normallyμ and Σ are unknown, but we can use estimated from a historical data set with n observations.

Phase I and Phase II : The application of Hotelling's T^2 statistic shall be categorized into two phases. Phase I tests whether the preliminary process was in control and phase II tests whether the future observation remains in-control (Alt,), (1). Phase I operation refers to the construction ofin-control data set. Same idea as Shewhart control chart, control limits are estimated from a period of in-control data. To obtain this in-control data, the raw data set needs to be purged. For instance, the outliers need to be removed and the missing data needs to be substituted with an estimate. During phase I

Operation, Hotelling'sT^2 statistic is calculated for each measurement and compared to the control limit, which will follows Chi-square distribution (according to Richard, A.J. & Dean, W.W.),(4)

$T^2 = (X - X)S^{-1\prime}(X - X) \sim X^2_{\alpha,p}$(Chi - square distribution) (1)

Also other research shows that the control limit follows Beta distribution (Mason, Young & Tracy,).(3)

$T^2 = (X - X)S^{-1\prime}(X - X) \sim \frac{(n-1)^2}{2}\beta\left(\alpha, \frac{P}{2}, \frac{n-p-1}{2}\right)$ (2)

n: number of preliminary observations

Both control limits will be approximate when the number of observations is large. The control limit based on Chi-square distribution is established on the assumption that are true values μ and Σ, which is just an approximate situation (Mason, Young & Tracy,) (3). Beta distribution is more precise and is a recommendable choice. After purging the raw data with Hotelling's T^2statistic, the in-control data set is ready for monitoring future observations which is termed as phase II operation. The control limit for determining future observation is different from the one in phase I. It follows an F distribution with *p* and (*n-p*) degrees of freedom. The idea of using Hotelling's T^2 statistic in phase I and phase II is the same. Each measurement is examined whether it is out-of-control by checking if it deviates extraordinarily from its sample mean. It should be reminded to choose the correct upper control limit on different purposes. The T^2 statistic calculated from all the observation will be plotted in a chart against observation serious and compared to the upper control limit. It should be noticed that there is no center line and the lower control limit is set to zero, because the meaning of T^2 statistic is a generalized distance between the observation and its sample mean.

II. CASE DEMONSTRATION

Starting by selecting any one of the *p* variables, followed by any of the (*p*-1) remaining Variables to condition on the first selected variable, the next step is to select any of the remaining (*p*-2) variables to condition on the first two selected variables. Iterating the same procedure will generate all the decomposition equations which give the same overall T^2 statistic. For instance, taking the first p variables, assuming we select p=1 is selected (remaining 2nd, and 3rd variables). The decomposition results are presented as follows;

$T_1^2 + T_{2.1}^2 + T_{3.12}^2$

$T_1^2 + T_{3.1}^2 + T_{2.1\ 3}^2$

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 in the above decompositions). Then, any of the remaining (p-2) variables not used in the first decomposition to be condition on the first two kept variables is selected. Computing in the same procedure will generate all the decomposition equations which produce the same overall T^2statistic. Hence, the complete T^2decomposition results using three variables (*p* = 3) are presented in equation (5). Hotelling's T^2 statistic helps us to determine whether a measurement isin-control or not.

Description of the methodology used: To reach the main purpose of this research, the data were generated from Wapco plc. The data is then used to generate a real data set collected at every point where there is a variation; three variables were monitored for twenty five observations were recorded.

The MYT Decomposition : The T^2 statistic can be decomposed into P-orthogonal components (Mason, Tracy and Young), (3) for instance, if you have P-variates to decompose, the procedure is as follows:

$T^2 = T_1^2 + T_{2.1}^2 + --- + T^2_{P.1,2,-------P-1}$ (3)

The first term is an unconditionalT^2, decomposing it as the first variable of the

$T_1^2 = \frac{(X_1 - X_1)^2}{S_1^2} = \frac{(-18)^2}{54} = 6(4)$

Where X_1 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

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

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

$$T^2 = \begin{cases} T_1^2 + T_{2.1}^2 + T_{3.12}^2 \\ T_1^2 + T_{3.1}^2 + T_{2.13}^2 \\ T_2^2 + T_{1.2}^2 + T_{3.12}^2 \quad (5) \\ T_2^2 + T_{3.2}^2 + T_{1.23}^2 \\ T_3^2 + T_{1.3}^2 + T_{2.13}^2 \\ T_3^2 + T_{2.3}^2 + T_{1.23}^2 \end{cases}$$

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

III. ILLUSTRATION:

In this segment, 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 Portland Cement Company in Lagos, Nigeria. The data consistsof temperature from a boiler machine inwhich twenty five observations were taken under different temperatures. The mean and covariance matrix given below was obtained from the historical data set.

$$s = \begin{pmatrix} 54 & 0.958 & 20.583 \\ 0.958 & 4.84 & 2.963 \\ 20.583 & 2.963 & 22.993 \end{pmatrix}, \quad \mathrm{X}= \begin{pmatrix} 525 \\ 513.56 \\ 538.92 \end{pmatrix}$$

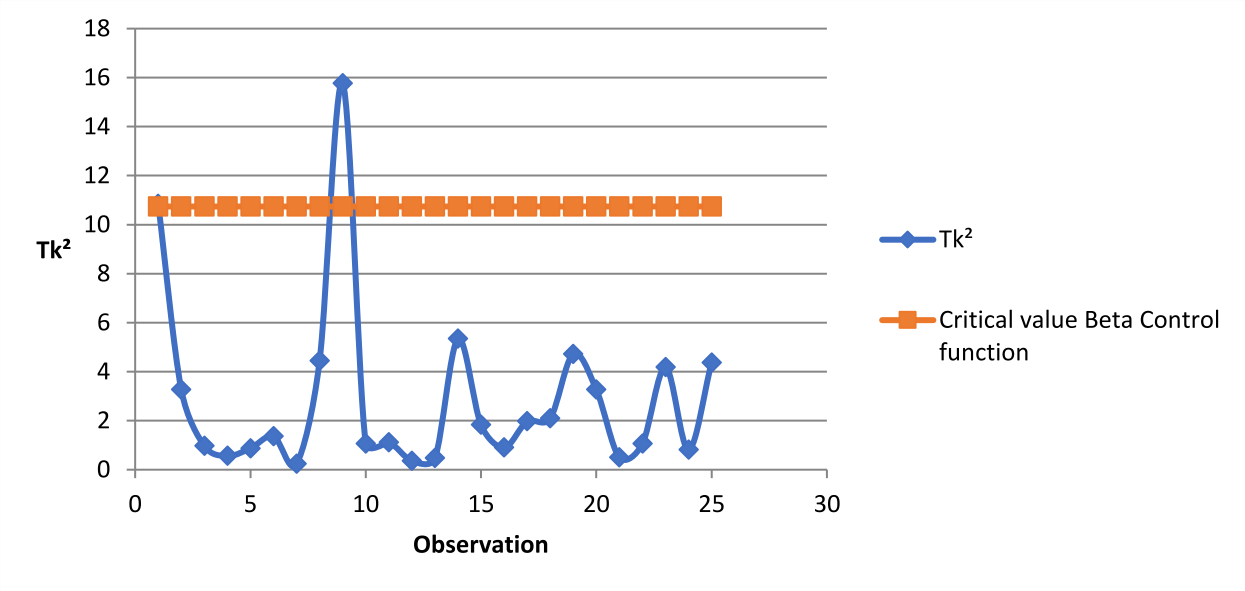
Having obtained the covariance matrix, the computation of T^2 statistic was done using STATGRAPHICS computer package and the values obtained for each of the twenty five observations are as shown in the table below.

TABLE 1:
Computation of Hotelling's T^2 Statistic for individual observation

No of observation	T_k^2	Critical value Beta Control function
1	10.85 * *	10.74
2	3.27	
3	0.97	
4	0.56	
5	0.87	
6	1.36	
7	0.24	
8	4.45	
9	15.77* *	
10	1.06	
11	1.12	
12	0.36	
13	0.47	
14	5.34	
15	1.83	
16	0.90	
17	1.97	
18	2.09	
19	4.72	
20	3.27	
21	0.50	
22	1.06	
23	4.18	
24	0.81	
25	4.37	

Below is the graph of Hotelling's T^2 statistic 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: Chartof Hotelling's T²Statistic



As it can be seen from the graph as well as the computation ofT^2 statistic in the table above, clearly it's shows that observation (1) and (9) are out of Control, as such, in order to detect the assignable signal, we consider observation(1), and the procedure is as follows:

Given that $X'(507, \quad 516, \quad 527)$and $\overline{X} = \begin{pmatrix} 525 \\ 513.36 \\ 538.92 \end{pmatrix}$

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

$$T^2 = (-18, 2.44, -11.92)\begin{pmatrix} 0.029 & 0.0109 & -0.0270 \\ 0.0109 & 0.228 & -0.039 \\ -0.0270 & -0.039 & 0.073 \end{pmatrix}\begin{pmatrix} -18 \\ 2.44 \\ -11.92 \end{pmatrix} = 10.85$$

UCL$= \frac{(n-1)^2}{n} \beta\left(\frac{\alpha}{2}, \frac{P}{2}, \frac{n-p-1}{2}\right) = 10.74$, where, α=0.005

Since T^2overall is significant, then we decompose the T^2 into its orthogonal components as to be able to find which among the set of variables contribute most to the out of control signals. Referred to the equation (5) above, where p =3 we had 18 components of $T^2$with 6 different equations each is yielding the same overallT^2.The calculation for each component in (4) as well as the sum of each term is as follows**:**

Starting with unconditional terms as given below

$T_{1,}^2 = \frac{(X_1 - X_2)^2}{S_1^2} = \frac{(-18)^2}{54} = 6$, $T_2^2 = \frac{(X_2 - X_2)^2}{S_2^2} = \frac{(2.64)^2}{4.84} = 1.44$, and$T_3^2 = \frac{(X_3 - X_3)^2}{S_3^2} = \frac{(-11.92)^2}{22.993} = 6.18$

Therefore, the UCL for unconditional terms is computed as follows:

UCL $= \frac{n+1}{n} . F_{\propto, 1, n-1} = 4.43$

It is show that only the variableT_2^2 is in Control since it is less than the UCL of the unconditional terms while $T_1^2$and T_3^2 are out of control, meaning that the variables X_1 andX_3 are part of signal contained in the observation vector.

To compute the conditional terms we proceed as follows:

$$T^2_{3.1,2} = T^2(x_1, x_2, x_3) - T^2(x_1, x_2)$$

To obtain $T^2(x_1, x_2)$ the original estimates of the mean vector and covariance structure is partioned to obtain the mean vector and covariance matrix of the sub vector.

$X^{(2)} = (x_1, x_2)$,the corresponding partitions are given as:

$$S_{22} = \begin{pmatrix} 54 & 0.958 \\ 0.958 & 4.84 \end{pmatrix}, \ \overline{X}^{(2)} = \begin{pmatrix} 525 \\ 513.36 \end{pmatrix}$$ The computation of,

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

$T^2(x_1, x_2)$ =$(-18,\ 2.64) \begin{pmatrix} 54 & 0.958 \\ 0.958 & 4.84 \end{pmatrix}^{-1} \begin{pmatrix} -18 \\ 2.64 \end{pmatrix} = 6.63.$ Thus,

$T^2_{2.1} = T^2(x_1, x_2) - T_1^2$ =6.63-6=0.63 and

$T^2_{3.1,2} = T^2(x_1, x_2, x_3) - T^2(x_1, x_2)$ =10.85-6.63=4.22

Also from the above, we have:

$T^2(x_1, x_2, x_3) = T_1^2 + T^2_{2.1} + T^2_{3.1,2}$ =6+0.63+4.22=10.85

For the fact that the first and third terms have higher values implies that the signal is contained in the first and third term $T^2_{3.1,2}$ .Next, to check whether there is signal in (x_1, x_2), separate the original observation into two groups as (x_1, x_2) and x_3 .

Having computed the value of $T^2(x_1, x_2)$ =6.63, thenCompare it to the

$$UCL(x_1, x_2) = \frac{2(25+1)(25-1)}{25(25-2)} F_{0.05}, 2, 23 = \frac{1248}{575}(3.42) = 7.4229$$

It is 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 first and third component.

It is a one possible decomposition, which was shown in order to illustrate a computing technique for the decomposition terms. We are guaranteed that, whichever decomposition we pick, we must obtain the same overall T^2 as follows:

$T_1^2 + T_{2.1}^2 + T_{3.12}^2$ = 6 + 0.63 +4.22 =10.85

$T_1^2 + T_{3.1}^2 + T_{2.1\ 3}^2$ =6 + 1.54 + 3.31 = 10.85

$T_2^2 + T_{1.2}^2 + T_{3.12}^2 = 1.44 + 5.19 + 4.22 = 10.85$

$T_2^2 + T_{3.2}^2 + T_{1.23}^2$ =1.44 + 6.83 + 2.58 = 10.85

$T_3^2 + T_{1.3}^2 + T_{2.13}^2$ = 6.18 + 1.36 + 3.31 = 10.85

$T_3^2 + T_{2.3}^2 + T_{1.23}^2 = 6.18 + 2.09 + 2.58 = 10.85$

With the aids of the result decomposed above; the value of each term of the decomposition was compared to the respective critical value as indicated in the table below:

Table: 2 MYT Decomposition Component for Observation (1)

Component	Value	Critical value
T_1^2	6* *	4.43
T_2^2	1.44	4.43
T_3^2	6.18* *	4.43
$T_{1.2}^2$	5.96	7.42
$T_{1.3}^2$	1.36	7.42
$T_{2.1}^2$	0.63	7.42
$T_{3.1}^2$	1.54	7.42
$T_{2.3}^2$	2.05	7.42
$T_{3.2}^2$	6.83	10.38
$T_{1.23}^2$	2.58	10.38
$T_{2.13}^2$	3.31	10.38
$T_{3.12}^2$	4.22	10.38

From the above table, we discovered that only two unconditionalT_1^2 and T_3^2 have significant values, which means there is a problem in $X_1$and X_3,we may conclude that X_1 and $X_3$are potential causes of the shifts. To verify the problems, we remove $T_1^2$and T_3^2 from the vector of observation (1) and check whether the sub vector signals or not.

T^2-$T_{1.23}^2 = 10.85 - 6 - 6.18 = -1.33 < 10.74.$

The outcomes Showthat, the sub vector is insignificant. The meaning of significant value of $T_{1.23}^2$ is that $X_1$and X_3 deviates from the variables relation pattern established from the historical data set.

IV. CONCLUSION

In this article, we were able to point out how Hotelling'sT^2 Statistic decomposition Procedure was employed to detect assignable signal .The T^2 decomposition is chosen in this research work due its ability to locate the potential cause of out-of-control signals. The use of multivariate control chart is very effective than the univariate control chart, although the interpretation of out-of-control signal in a multivariate control chart is more complex than the univariate control chart. Thus, the T^2 decomposition is being derived and applied in this research. This is to encourage quality control staffs on the significance of theT^2 decomposition in detecting variable(s) that significantly contribute to out-of-control signal. Broadly speaking, Control Charts are used so as to be able to distinguish between assignable and natural causes in the variability of quality goods produce. With regards to this, verifying which combination of quality characteristics is responsible for the signal. There fore the Control Chart plays a key role to inform us the appropriate next line of action to be taken in order to enhance the quality.

REFRENCES:

1. Alt, F.B., (1985), Multivariate quality control, *Encyclopedia of the statistical sciences*, Vol.6, Pp.111-122
2. Hotelling, H., (1931).Thegeneralization of student's ratio. *The Annuals of Mathematical Statistics*v.2, pp.360-378.
3. Mason, R.L., Tracy, N.D. & Young, J.C., (1992), Multivariate control charts for individual observations,*Journal of Quality Technology*, Vol.24, No.2, pp.88-95
4. Richard, A.J. & Dean, W.W., (2002).*Applied multivariate statistical analysis*, N.J.: Prentice Hall

APPENDIX I

Boiler Temperature Data

N0.of observation	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	542
7	522	513	537
8	527	509	537
9	533	514	528
10	530	512	528
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