Proceedings of the 2013 International Conference on Information, Operations Management and Statistics (ICIOMS2013), Kuala Lumpur, Malaysia, September 1-3, 2013

Enhancing Hotelling T^2 Control Chart Performance with Decile and Trimmed Means

Herma Mohd Hanif [a], Norazan M. R [b], Khairul A. Rasmani [c], Nor Azura M. G [d]

[a,b,c,d] *Faculty of Computer and Mathematical Science*
Universiti Teknologi MARA Shah Alam
40450 Selangor Darul Ehsan
MALAYSIA

Abstract

Hotelling T^2 control chart is a multivariate control chart that takes multiple characteristics into consideration at a time. However, the performance of the classical Hotelling T^2 control chart which uses classical mean in phase one is easily affected if the data set contain outliers and consequently the upper control limit in phase two would also be affected. Thus, in this paper, decile and trimmed means were used to replace the ordinary mean in phase one to improve the efficiency of the classical Hotelling T^2 control chart. We investigated the performance of the proposed method through a simulation study and numerical example based on higher probability of detection which indicates good performance. Simulation results showed that the performance of Hotelling T^2 control chart with a trimmed mean in phase one consistently outperformed the ones with decile and ordinary means.

Keywords: Hotelling T^2 control chart; Trimmed mean; Decile mean

---

1. Introduction

Individual Hotelling T^2 control chart is widely used in multivariate chart. Harold Hotelling introduced the generalized Student- t statistic and control procedure based on new charting statistics. Hotelling T^2 control chart is named after Harold Hotelling as an honor to him. Many studies have been conducted on Hotelling T^2 control chart among researchers like in Deo Kumar, Mandana, Shoaja'eddin and Syed Yahya [1-4] that study on Hotelling T^2 control chart.

Control charts can be used when the data are assumed to be normal and free of outliers. However, this is rarely true in practice. The presence of outliers will affect the accuracy of the control chart in detecting abnormalities. Data used in phase one must not consist any outliers in order to get a good control limit for phase two. Screening data and removing outliers are not sufficient enough. The implementation of robust technique in individual classical Hotelling T^2 control chart helps to overcome the effect of outliers' in phase one and phase two of quality control process. Hotelling T^2 control charts with robust means are expected to produce more efficient result in detecting process that is out of control. Trimmed mean is a type of robust mean and a method of averaging that removes a small percentage of the largest and smallest values before calculating the mean. After removing the specified observations, the trimmed mean is found using an arithmetic averaging formula. It is also being used by researchers such as Kumar and Mudholkar that implement the method in Hotelling T^2 control chart [1] and also by Term [5]. The benefit of trimmed mean is that the standard errors associated with them will be smaller than it will be for the untrimmed mean. Trimmed mean in univariate cases is straightforward compared to multivariate [1]. Multivariate setting considers coordinate by coordinate trimming of the multivariate sample. To find a trimmed mean, the x% largest and smallest scores are deleted and the mean is computed using the remaining scores. This study used 10% of trimmed which can be conducted as follows:

Step 1: Sort the data in ascending order.
Step 2: Discard selected percentage of higher and lower values in data.
Step 3: Recompute the mean with the remaining data.

Meanwhile decile mean is another type of measures of central tendency since it is based on deciles. This measure should be fairly robust as it automatically discards extreme observations or outliers from both tails but at the same time is more informative than the median or interquartile mean [6]. It is less sensitive to large values as compared to other existing measures.It depends on the eighty percent of a sample, a population, or a probability distribution and automatically discard extreme observations or outliers from both tails. A decile is any of the nine values that divide the sorted data into ten equal parts, so that each part represents 1/10 of the sample or population for raw data as well as decile from the frequency distribution [6]. It is the summation of all deciles divided by the number of deciles [6]. Thus, the i^{th} decile for a sample with size N can be computed as follows:

$$\text{Decile}_i = \text{size of } \left[\frac{i\,(N+1))}{10}\right]^{th} \text{ item of the series} \quad \text{for } i = 1,2,....,9 \tag{1}$$

2. Methodology

This section presents the methodology used in this study. The first section explains in details the steps involved in computing the classical Hotelling T^2 control chart. This is followed by explanation on the integration of decile and trimmed in phase one of Hotelling T^2 control chart. The resulting Hotelling T^2 control charts with decile and trimmed mean will be termed respectively as Hotelling T^2 (decile) and Hotelling T^2 (trim).

2.1 Hotelling T^2 Control Chart

Hotelling T^2 control chart consists of two phases. Phase one of control chart is where the data or the process assumed to be in control or normally distributed. Meanwhile the second phase, phase two is where the true data is applied in the chart to detect any abnormalities of the process. Below is the step by step approach for conducting classical Hotelling T^2 control chart. Note that in this control chart, the Upper Control Limit (UCL) is the main interest.

Phase One

Step 1 : Decide on sample size n, number of variables p, and confidence level (1- α)

Step 2 : Collect the Phase one data ($x_1,x_2..x_n$) at well-defined periodic intervals

Step 3 : Use the Phase one data, compute the mean, $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ and the covariance, $s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$ to get

$$T^2(t) = (x_i - \bar{x})\, s^{-1}\, (x_i - \bar{x})^t \tag{2}$$

Step 4 : Define the outliers by using UCL based on Beta distribution

$$\text{UCL} \sim \left[\frac{(n-1)^2}{n}\right] B_{(p/2,\, n-p-1/2)} \tag{3}$$

Step 5 : Remove the observations which are considers as outliers (if data exceed UCL).

Step 6 : Estimate new $\bar{x}$ and s using sample without outliers (as obtained in Step 5 above).

Phase Two

Step 1 : Compute $T^2(g) = (x_g - \bar{x})\, s^{-1}\, (x_g - \bar{x})^t$ using sample without outliers (as obtained in Step 5 in Phase one).

Step 2 : Compute UCL using F distribution;

$$\text{UCL} \sim \left[\frac{p(n+1)(n-1)}{n(n-p)}\right] F_{(p,n-p)} \tag{4}$$

Step 3 : Interpret the chart and look for out-of-control points or patterns. Diagnose the process if needed.

2.2 *Hotelling T^2 Control Charts with Decile and Trimmed Means in Phase One*

To enhance the performance of the classical Hotelling T^2 control chart, decile mean will be used in phase one to replace ordinary mean, $\bar{x}$. The following procedures explain the details:

Phase One

Step 1 : Decide on sample size n, number of variables p, and confidence level (1- α)

Step 2 : Collect the Phase one data ($x_1, x_2 .. x_n$) at well-defined periodic intervals

Step 3 : Use the Phase one data, compute decile mean, $\bar{x}_d$ and covariance based on decile mean, s_d . Compute Mahalanobis distance as follows:

$$T^2(t) = \left(x_i - \bar{x}_d\right) s_d^{-1} \left(x_i - \bar{x}_d\right)^t \quad (5)$$

Step 4 : Define the outliers by using UCL based on Beta distribution

$$\text{UCL} \sim \left[\frac{(n-1)^2}{n}\right] B_{(p/2,\ n-p-1/2)} \quad (6)$$

Step 5 : Remove the observations which are considered as outliers (data that exceed UCL)

Step 6 : Estimate new $\bar{x}$ and s using sample without outliers (obtained in Step 5 above).

Phase Two

Step 1 : Compute $T^2(g) = \left(x_g - \bar{x}_d\right) s_d^{-1} \left(x_g - \bar{x}_d\right)^t$ using sample without outliers (as obtained in Step 5 in Phase one).

Step 2 : Compute UCL using F distribution;

$$\text{UCL} \sim \left[\frac{p(n+1)(n-1)}{n(n-p)}\right] F_{(p,n-p)} \quad (7)$$

Step 3 : Interpret the chart and look for out-of-control points or patterns. Diagnose the process if needed.

The same steps are repeated to get Hotelling T^2 control charts with trimmed mean, $\bar{x}_t$, by simply replacing the ordinary mean with trimmed mean in phase one. The rest of the procedures remain the same.

3. A Simulation Study

The performance of the enhanced control charts with decile and trimmed means are assessed through a simulation study. The following simulation design describes the procedure of the simulation study. Table 1 summarizes the factors together with their respective various sizes and levels used in the simulation study.

Table 1. Table of Factors and Level of Factors

Factor	Level of Factors
sample sizes (n)	20,30,50,70,100,130,150,170, 200
number of variables (p)	2, 5, 10
percentage of outliers	20
Shift (mean)	3,5

A total of 1500 samples of size n from multivariate normal distributions,(MVN(0,I)) were generated for different combination of factors and levels. Each of the attributes are assumed to be independent with each other. The performance of the enhanced control charts is determined by the probability of detection. The best method is expected to produce the highest probability of detection. The following give the details of the calculation:

Step 1 : Generate data X of size n from multivariate normal distributions, i.e $X \sim MVN(0,I)$ based on the conditions given in Table 1.

Step 2 : Apply the Hotelling T^2 control chart to detect outliers.

Step 3 : Observe the number of the outliers, θ_1.

Step 4 : Repeat Step 1, Step 2, and Step 3 for 1500 times and we will get the number of outliers as $\theta_1, \theta_2, \ldots, \theta_{1500}$

The mean number of outliers for 1500 times is $\bar{\theta} = \frac{\sum_{i=1}^{n} \theta_i}{1500}$. Thus, the the probability of detection is obtained by dividing $\bar{\theta}$ with the sample size n.

3.1.1 Simulations Results

Table 2.(a), Table 2.(b) and Table 2.(c) below give results of the simulation study according to groups of variables and shifts.

Table 2. (a) Table Probability of Detection when data sets contain 20% outliers with two attributes

p	Shift	n	Hotelling T^2	Hotelling T^2 (trim)	Hotelling T^2 (decile)
2	3	50	0.13810	**0.17390**	0.13140
		70	0.14660	**0.17870**	0.14440
		100	0.15200	**0.18140**	0.15210
		130	0.15490	**0.18320**	0.15670
		150	0.15610	**0.18370**	0.15850
		170	0.15740	**0.18420**	0.16070
		200	0.15920	**0.18460**	0.16300
	5	50	0.15580	**0.18890**	0.14620
		70	0.16470	**0.19200**	0.15950
		100	0.16970	**0.19330**	0.16660
		130	0.17220	**0.19430**	0.17080
		150	0.17380	**0.19450**	0.17300
		170	0.17460	**0.19470**	0.17420
		200	0.17550	**0.19480**	0.17620

In Table 2. (a), the attributes used are two with a shift value of three and five. From the table given, it can be seen that Hotelling T^2 (trim) consistently gives the highest probability of detection as compared to Hotelling T^2 control chart with ordinary mean and decile mean. We also noticed that the probability of detection of the Hotelling T^2 (trim) becomes higher as the sample size increases. It seems that the performance of the Hotelling T^2 (trim) is not affected as the shift value gets higher. On the other hand, the ordinary and Hotelling T^2 (decile) gives almost the same values of detection probabilities. Meanwhile in Table 2.(b) and Table 2(c) show simulation results when the attributes are increased to five and ten respectively. Uniformly, the results are almost the same as in Table 2.(a) whereby the Hotelling T^2 (trim) gives the highest probability of detection, followed by the Hotelling T^2 (decile) and the Hotelling T^2 control chart using ordinary mean.

Table 2. (b) Table Probability of Signal when data sets contain 20% outliers with five attributes

p	Shift	n	Hotelling T^2	Hotelling T^2 (trim)	Hotelling T^2 (decile)
5	3	50	0.21160	**0.21270**	0.21240
		70	0.19930	**0.20000**	0.19930
		100	0.19950	**0.19990**	0.19960
		130	0.19970	**0.19990**	0.19970
		150	0.19970	**0.19990**	0.19980
		170	0.19980	**0.19990**	0.19980
		200	0.19980	**0.19990**	0.19990
	5	50	0.21160	**0.21270**	0.21240
		70	0.20000	**0.20000**	0.20000
		100	0.19998	**0.20000**	0.20000
		130	0.19998	**0.20000**	0.20000
		150	0.19999	**0.20000**	0.20000
		170	0.20000	**0.20000**	0.20000
		200	0.19998	**0.20000**	0.20000

Table 2. (c) Table Probability of Signal when data sets contain 20% outliers with ten attributes

p	Shift	n	Hotelling T^2	Hotelling T^2 (trim)	Hotelling T^2 (decile)
10	3	50	-	-	-
		70	0.21319	**0.35183**	-
		100	0.21747	**0.21369**	0.20714
		130	0.20000	**0.21833**	0.20935
		150	0.20000	**0.20000**	0.20000
		170	0.20000	**0.20000**	0.20000
		200	0.20000	**0.20000**	0.20000
	5	50	-	**-**	-
		70	0.21319	**0.27086**	-
		100	0.21801	**0.21369**	0.20714
		130	0.20000	**0.21833**	0.20935
		150	0.20000	**0.20000**	0.20000
		170	0.20000	**0.20000**	0.20000
		200	0.20000	**0.20000**	0.20000

Plots Fig. 1. summarizes the performance of the three control charts for simulated data sets with a number of variables two, and shift mean equals to three.We can conclude that when data sets are contaminated with 20% of outliers, the Hotelling T^2 (trim) control chart outperforms the rest with highest detection probabilities.

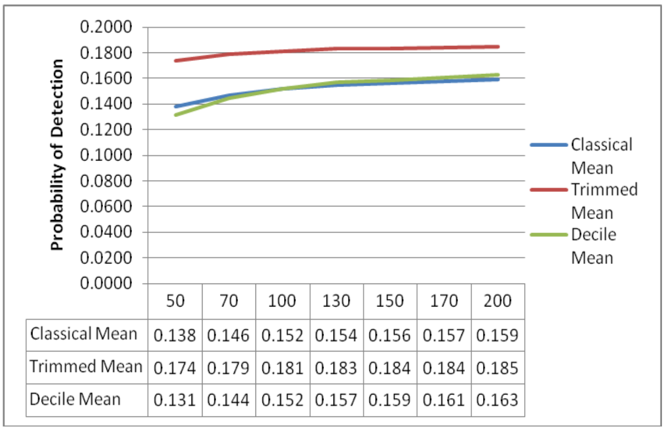

Fig. 1. Comparisons of Classical, Trimmed and Decile Hotelling T^2 control chart

3.2 An Application on Residential Water Consumption Data in Detecting Excessive Water Usage

To illustrate the usability of the proposed methods, we apply the method to residential water consumption data (in liter) to detect excessive water usage. The data represent the domestic water consumption in selected groups of study of 439 of consumers. Ten quality characteristics are considered in the study which are expected to contribute to the amount of domestic water consumption. The listed characteristics are; the family members, number of toilets, the number of times brusing teeth, number of times washing hands and face, number of times taking shower, number of times flushing toilet, number of times doing hand wash (clothes), number of times doing machine wash, number of times cooking, number of times doing car wash [7]. Fig. 2. shows the plot comparing the probability detection of the three Hotelling T^2 control charts. The plot shows that the classical Hotelling T^2 gives the lowest detection of outliers as compared to the other two control charts. We notice that the Hotelling T^2 (trim) and Hotelling T^2 (decile) control charts produce the approximately the same amount of probability of detection.

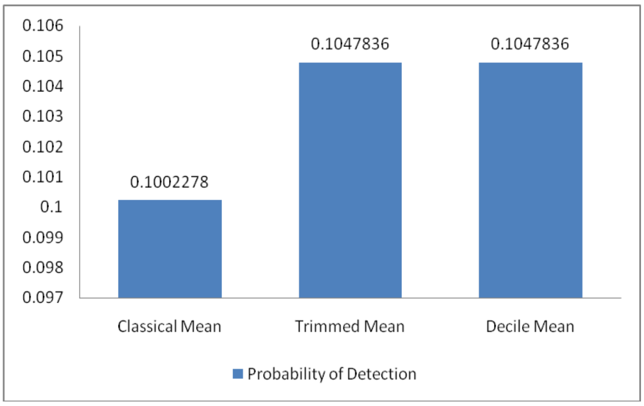

Fig. 2. Probability Detection of Classical, Trimmed and Decile Hotelling T^2 control chart

4. Conclusion

This paper gives a comparison study on the implementation of trimmed and decile means in Hotelling T^2 control chart . The results of the study presented in this paper show that the Hotelling T^2 (trim) and Hotelling T^2 (decile) can be used as an alternative to replace the classical Hotelling T^2 control chart when outliers are present in the data set. The probability of detection of the Hotelling T^2 control chart becomes higher as the sample size increases and as the number of variables increases. The illustration of the methods using residential water consumption data sets show that the robust trimmed and decile mean gives a uniform result and higher that the classical mean. It can be concluded that Hotelling T^2 with a trimmed mean outperform other methods.

Acknowledgments

This research work is funded by the Universiti Teknologi MARA under the Research Intensive Faculty Grant Scheme (RIF) with reference number 600-RMI/DANA 5/3/RIF (411/2012).

References

[1] Srivastava, D.K. and G. S.Mudholkar, *Trimmed T-Square : A Robust Analog of Hotelling's T-Square*. Journal of Statistical Planning and Inference, 2001. **97**: p. 343–358.

[2] Mohammadi, M., H. Midi, and J. Arasan. *Re-weighted Robust Control Charts for Individual Observations*. in *Proceedings of the 6th IMT-GT Conference on Mathematics, Statistics and its Applications*. 2010. Universiti Tunku Abdul Rahman, Kuala Lumpur, Malaysia.

[3] Chenouri, S.e., S.H. Steiner, and A.M. Variyath, *A Multivariate Robust Control Chart for Individual Observations*. Journal of Quality Technology, 2009. **41**(3).

[4] Yahaya, S., et al., *An Alternative Hotelling T-Square Control Chart Based on Minimum Vector Variance (MVV).* Modern Applied Science, 2011. **5**(4).

[5] Term, W., *Introduction to the Bootstrap and Robust Statistics.* 2009. **PSY711/712**.

[6] Rana, S., et al., *Decile Mean: A New Robust Measure of Central Tendency*. Cjiang Mai Journal Science, 2012. **39**(3): p. 478-485.

[7] Hanif, H.M., K.A. Rasmani, and N.M. Ramli. *Detecting Residential Household Water Leakage and Wastage using Individual Hotelling T-Square Control Chart*. in *IJAS American - Canadian International Conference*. 2013. Toronto, Canada.