EFFECT OF CORRELATION BETWEEN VARIABLES FOR MULTIVARIATE CONTROL CHARTS TO MONITOR THE PROCESS MEAN

R. A. B. Abeygunawardana[1] and R. De Silva[2]

[1]Department of Statistics, University of Colombo, Sri Lanka, rab_abey@stat.cmb.ac.lk
[2]Department of Statistics, University of Colombo, Sri Lanka, renu313@gmail.com

ABSTRACT

Control charts are one of the main techniques used to identify shift from target in quality characteristics. Depending on the number of quality characteristics to be monitored univariate or multivariate control charts can be applied. Monitoring correlated quality characteristics independently may be misleading due to correlation between variables. In this study average run lengths (ARL) of univariate ($\bar{X}$ chart) and multivariate (Hotelling's T^2) control charts were studied. Simulations were carried out for ARL calculation. Out-of-control ARLs were calculated by giving different size of mean shifts to simulated data. When the correlation between two variables is increasing, the both in-control and out-of-control ARL of multivariate control chart is found to be decreasing.

Keywords: Statistical Process Control, Control Chart, Univariate, Multivariate, ARL

1. INTRODUCTION

Statistical Process Control (SPC) plays a major role in monitoring and controlling quality characteristics of a manufacturing products and services. SPC uses statistical methods to identify any special cause of variation in a process. Control charts are one of the main techniques used to identifying shift in quality characteristics from its target value. Depending on the number of quality characteristics to be monitored control charts can be classified as univariate control charts and multivariate control charts. Traditionally control charts are monitored quality characteristics individually based on the assumption that they are independent and identically distributed. Univariate control charts are often used in detecting unusual changes in variables that are independent and thus are not affected by the behavior of other variables. In real world there may be situations in which more than one related quality characteristics should monitored simultaneously. For such a situation multivariate controls may more suitable than univariate charts. Variables of a multivariate process often interrelated and form correlated data sets. When the variables do not behave independently of one another, they must be examined together as a group, not separately.

Monitoring correlated quality characteristics independently by univariate control charts can be misleading due to correlation between variables. Multivariate control charts are more efficient than the simultaneous operation of several univariate control charts. While Shewhart $\bar{X}$-chart is the most commonly used univariate chart to monitor the process mean Hotelling's T^2 chart is commonly used multivariate control chart as an alternative to the univariate Shewhart control charts. Upper and lower control limits of $\bar{X}$-chart are $\bar{\bar{X}} \pm L\hat{\sigma}$ with the central line $\bar{\bar{X}}$. Here L is the distance of the control limit from the center line (Montgomery, 1995). In the traditional control charts it is assumed that Measurements are independent of each other and the mean of the measurements are normally distributed. Hotelling's T^2 chart is formed by plotting Hotelling's T^2 statistics vs sample number. Bersimis, Psarakis and Panaretos (2006) have discussed the basic procedures for the implementation of multivariate statistical process control via control charting. Bersimis et al (2006) have cited that Lowry and Montgomery (1995) have suggested different categories for test statistic of Hotelling's T^2 chart working with individual observations and working with rational subgroups. When constructing Hotelling's T^2 chart, it is used usual sample mean vector and variance-covariance matrix to calculate the T^2 statistics. But it performs poorly. Minimum volume ellipsoid (MVE) and minimum covariance determinant (MCD) are alternative estimators for construct the T^2 chart. Yanez, Gonzalez and Vargas (2009) have proposed Hotelling's T^2 chart using the biweight S estimators for the location and dispersion parameters when monitoring multivariate individual observations.

The following three forms are applicable to determine the distribution of Hotelling's T^2 statistic.

1. Suppose that μ and $\sum$ of the underlying multivariate normal (MVN) distribution are known. Then the distribution of the T^2 statistics is given by,

$$T^2 = (X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi_p^2$$

Here χ_p^2 represents a chi-square distribution with p degree of freedom. The T^2 distribution depends only on p, which is the number of variables in the observation vector X.

2. Suppose that μ and $\sum$ of the underlying MVN distribution are unknown and are estimated using estimators $\bar{X}$ and sample standard deviation (S). Then the distribution of T^2 statistics is given by as

$$T^2 = (X-\bar{X})'S^{-1}(X-\bar{X}) \sim \left[\frac{p(m+1)(m-1)}{m(m-p)}\right] F_{(p,m-p)}$$

Where $F_{(p,m-p)}$ is a F distribution with p and (n-p) degree of freedom. This distribution depends on the number of samples as well as on the number of variables being examined.

3. Suppose that the observation vector X is not independent of the estimators $\bar{X}$ and S but is included in their computation. In this situation the distribution of T^2 statistic is given as

$$T^2 = (X-\bar{X})'S^{-1}(X-\bar{X}) \sim \left[\frac{(m-1)^2}{m}\right] B_{(\frac{p}{2},\frac{m-p-1}{2})}$$

Where $B_{(\frac{p}{2},\frac{m-p-1}{2})}$ represents a beta distribution with parameters p/2 and (m-p-1)/2. This distribution depends on the number of variables, p, and number of samples m.

Suppose that m samples, each of size n>1, are available and that p is the number of quality characteristics observed in each sample. Now suppose that S is used to estimate $\sum$ and $\bar{\bar{X}}$ is used to estimate μ. Then Hotelling's T^2 statistic is,

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

For a given α level control limits of is determined by

$$UCL = \left[\frac{p(m-1)(n-1)}{mn-m-p+1}\right] F_{[\alpha,p,mn-m-p+1]}$$

$$LCL = 0$$

When the parameters are known, $\chi^2_{(\alpha,p)}$ is used as the *UCL*. This study discusses the details of only constructing Hotelling's T^2 for subgroup data and its performance.

Young and Mason (2002) have discussed the details of control limits determinations of the Hotelling's T^2 chart and distribution of the T^2 statistic.

One of the main objectives of this research is to study the Average Run Length (ARL) performance of univariate X-bar chart and Hotelling's T^2 chart constructed for variables with different correlations.

2. METHODOLOGY

The main assumptions of the univariate control charts are, data should be independent and normally distributed. Since it considered more than two variables in the multivariate context data are correlated. Positively correlated and normally distributed two variables were generated using developed R code in this study. First of all 1000 pairs of observations with a specific correlation were generated. Following procedure was carried out for simulation.

Step 01 : Generate 1000 random observations which are normally distributed with given mean and standard deviation.

Step 02 : Generate another 1000 normally distributed random numbers, as correlation is to be a specific value with previous data. The mean and the standard deviation of new 1000 observations are unknown.

Then another 1000 such a pairs of observations were generated with the same correlation. First set of observations of each pair and second set of observations of each pair were average separately to get subgroup of size two. It was found that the correlation between averaged variables were approximately the same as original variables.

To calculate in-control ARL 100 simulations were carried out and 2000 such data sets were simulated to calculate out-of-control ARL. Correlations between two variables were set to 0.25, 0.5, 0.75 and 0.95 in this study. Mean shifts were given to standard normally distributed data in order to calculate out-of-control ARL. Here five levels of mean shifts in term of standard deviation units such as 0.5σ, 1σ, 1.5σ, 2σ and 3σ were given to calculate out-of-control ARL. Simulation was carried out for ARL calculation.

3. RESULTS

The in-control ARL of $\bar{X}$ chart was set to 350. The in-control ARL of the Hoteling's T^2 chart identified as approximately 130.

Table 1: ARL performance of Hoteling's T^2 chart at various levels of r

Mean shift	r = 0.25	r = 0.5	r = 0.75	r = 0.95
0.0	129.28	131.34	136.46	140.21
0.5	89.31	83.41	79.8	78.62
1.0	76.34	71.71	67.56	65.8
1.5	66.93	65.32	64.86	64.35
2.0	60.21	59.76	59.5	58.9
3.0	56.86	56.34	55.93	55.91

It was found that when the mean shift is increasing out-of-control ARL of Hoteling's T^2 chart is decreased. The decreasing rate of out-of-control ARLs was decreased when the correlation is increasing. And when the correlation is increasing, it can be seen that both in-control and out-of-control the ARL of Hoteling's T^2 is decreasing.

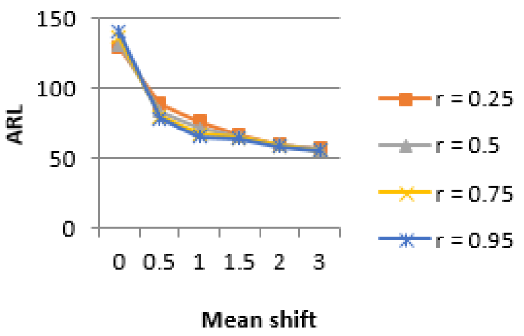

Figure 1: ARL performance of Hoteling's T^2 chart at various level of r

Table 2 shows that, when the mean shifts are increasing the out-of-control ARL of $\bar{X}$ chart is decreasing. The correlation between two variables did not severely affected to the in-control ARL and out-of-control ARLs of $\bar{X}$ chart when the shit size is large. When comparing two $\bar{X}$ charts and corresponding Hotelling's T^2 chart in in-control state it was found that the Hoteling's T^2 chart performed better than two individual $\bar{X}$ charts. Also this result is valid for small size shifts in the process mean. At the large shifts, $\bar{X}$ charts detected signal earlier than Hotelling's T^2 chart.

Table 2: ARL performance of X bar chart and Hoteling's T^2 chart at various level of correlation

Mean shift	Correlation = 0.25			Correlation = 0.5			Correlation = 0.75			Correlation = 0.95		
	X bar chart for first variable	X bar chart for second variable	Hotelling's Tsquare	X bar chart for first variable	X bar chart for second variable	Hotelling's Tsquare	X bar chart for first variable	X bar chart for second variable	Hotelling's Tsquare	X bar chart for first variable	X bar chart for second variable	Hotelling's Tsquare
0	351.59	350.12	129.28	339.16	341.62	131.34	340.13	344.05	136.46	336.68	342.41	140.21
0.5	131.44	136.25	89.31	141.85	136.7	83.41	147.26	144.32	79.8	148.03	147.56	78.62
1	43.46	36.59	76.34	39.5	36.09	71.71	41.67	39.45	67.56	41.45	41.43	65.8
1.5	12.73	12.96	66.93	10.72	11.59	65.32	11.72	12.99	64.86	14.08	13.67	64.35
2	5.97	5.31	60.21	5.08	5.66	59.76	6.01	5.89	59.5	6.02	5.21	58.9
3	1.51	1.26	56.86	1.38	1.41	56.34	1.49	1.21	55.93	1.36	1.32	55.91

When the in-control ARLs of the $\bar{X}$ charts are set to be 350 then the in-control ARLs of Hotelling's T^2 charts are approximately equal to 130. When the correlations increase, in-control ARLs of Hotelling's T^2 chart are increasing. But correlation does not affect for the in-control ARLs of individual $\bar{X}$ charts. Also when mean shift is given, the out-of-control ARLs decrease in both $\bar{X}$ chart and Hotelling's T^2 chart. But the decreasing rate of $\bar{X}$ chart is higher than decreasing rate of Hotelling's T^2 chart. The ARLs of the Hotelling's T^2 chart show lower values than ARLs of individual two $\bar{X}$ charts at the mean shift 0.5.

That means Hotelling's T^2 chart is quickly detect the out of control signal than two of individual univariate $\bar{X}$ charts. The ARLs of Hotelling's T^2 chart decreases while correlation range increase, from 0.25 to 0.95. When increase the correlation the performance of Hotelling's T^2 chart is higher than corresponding individual $\bar{X}$ charts. Although in large shifts (1σ to 3σ), the ARL of $\bar{X}$ charts is lower than ARL of Hotelling's T^2 chart.

That means individual $\bar{X}$ charts detect out of control signal earlier than Hotelling's T2 chart in large mean shifts. This is same, when increasing the correlation. Finally it can be said that Hotelling's T2 chart is best when comparing corresponding two $\bar{X}$ charts for detecting the out of signal at low level of mean shifts at any levels of correlation. Also it is better to use two $\bar{X}$ charts when comparing corresponding Hotelling's T2 chart at large mean shifts.

4. CONCLUSION

The correlation of two correlated data sets was approximately same even for the standardized variables. The out-of-control ARL of $\bar{X}$ chart is decreased when the mean shift is increasing. The
correlation between two variables did not severely affect to the in-control ARL of $\bar{X}$ chart. When the in-control ARL of $\bar{X}$ chart is fixed to 350 then the in-control ARL of the Hotelling's T^2 chart is approximately 130. When the mean shift is increasing out-of-control ARL of Hotelling's T^2 chart is decreasing. But that decreasing rate is decreased when increasing the mean shifts. Also when increasing the correlation, it can be seen that the ARL of both in-control and out-of-control of Hotelling's T^2 chart decrease. The decreasing rate of out-of-control ARLs was decreased even increase the correlation.

When comparing the two $\bar{X}$ charts and corresponding Hotelling's T^2 chart, Hotelling's T^2 chart performed better than two individual $\bar{X}$ charts at process was in-control. This result is valid for small size of mean shifts in the process. At the large size of shifts, $\bar{X}$ charts issue a signal earlier than Hotelling's T^2 chart. ARL values are decreased while the mean shifts increase from 0.5 to 3.When mean shift is large, the decreasing rate of ARL values are increased.

5. REFERENCES

[1] S. Bersimis, S. Psarakis, & J. Panaretos, "*Multivariate Statistical Process Control Charts: An Overview.* Quality and Reliability Engineering International", 23, pp.517-543, 2006.

[2] C. A. Lowry, and D. C. Montgomery, "*A review of multivariate control charts"* IIE Transactions, volume 27, pp. 800-810, 1995.

[3] D. C. Montgomery, "*Introduction to Statistical process Control* (6th ed.), New York: Wiley, 2006.

[4] Z. Wu, S. & Haridy "*Univariate and multivariate control charts for monitoring dynamic-behavior processes: a case study"*, Journal of Industrial Engineering and Management. 2(3), pp.464-498, 2009.

[5] S. Yanez, N. Gonzalez and N. Vargas N., "Hotelling's T^2 control chart based on robust estimators, *Dyna rev.fac.nac.minas"* [online]. vol.77, n.163, pp. 239-247. ISSN 0012-7353, 2009.

[6] J. C. Young amd R. L. Mason, "*Multivariate Statistical Process Control with Industrial Applications"* SIAM, ISBN 0898714966, 2002