HOW TO CONTROL PROCESS VARIABILITY MORE EFFECTIVELY: THE CASE OF A B-COMPLEX VITAMIN PRODUCTION PROCESS

M.A. Djauhari[1] and **I. Mohamad[2]**

[1, 2]Department of Mathematics
Universiti Teknologi Malaysia, Malaysia
maman@utm.my, im@mel.fs.utm.my

ABSTRACT

This paper presents a new procedure for controlling process variability where (i) the quality of the process is determined by more than one quality characteristic, and (ii) the correlations among those quality characteristics must be taken into consideration. The ability of this procedure to detect an abrupt shift of covariance structure is an improvement on the standard practice based on generalised variance. An experience of the production process of the B-complex vitamin is reported to illustrate the merit of the proposed procedure.

OPSOMMING

Die metodes van 'n nuttige gehaltebeheerprosedure vir die bepaling van gehaltebeheerverlies wat 'n chemiese produk met meervoudige gehalte-eienskappe besit, word beskryf teen die agtergrond van kovariasie. Die prosedure word getoets in die praktyk by die vervaardiging van 'n gegewe chemiese produktipe. Aspekte van voordele en nadele van die metode word ook gewetensgetrou behandel.

1. INTRODUCTION

'Reduce process variability' is the basic philosophy used in any manufacturing industry to improve the quality of the process and its products. It is common practice to materialise that philosophy and to visualise the history of process variability using a control chart. This chart is one of the 'magnificent seven' tools in all quality improvement initiatives, alongside the histogram, check sheet, cause-and-effect diagram, Pareto chart, scatter plot, and stratification. (See, for example, Rooney et al. [6] for further discussion of these tools.) All of them can be found in any standard book on statistical quality control. Nowadays, of these tools only control charting is a dynamic research area, especially in multivariate process control. By 'multivariate process' is meant a process where (i) its quality is determined by more than one quality characteristic, and (ii) the correlations between such quality characteristics must be taken into consideration.

This paper deals with Shewhart type control charting for detecting abrupt variability change in a multivariate normal process. The process refers to any multivariate process where all the quality characteristics together follow a multivariate normal distribution, a distribution model often used in this setting. Statistically, the variability of this process is represented in the form of a rectangular covariance matrix. It is a symmetric matrix of p rows and p columns where (i) p is the number of quality characteristics, (ii) the k-th diagonal element is the variance of the k-th quality characteristic and (iii) the element at the i-th row and j-th column is the covariance of the i-th and j-th quality characteristics.

Many Shewhart type control charting procedures are available for detecting abrupt change in multivariate normal process variability. One of the most widely-used control charts is the generalised variance-based chart, which can be found in the literature on multivariate process control. For certain recent developments of the generalised variance-based control charting procedure, see Mason et al. [5], who eliminate the condition that the number of observations must be greater than the number of quality characteristics in Phase II control charting. See also Djauhari [3] who presents a method for removing the bias in the control limits of the generalised variance chart and Alt and Smith [1] for further generalised variance-based control charting procedures. However, careful attention is needed when using a generalised variance control chart or a generalised variance-based control chart, since generalised variance – which is algebraically defined as the determinant of the covariance matrix – is only a scalar simplification of the complex covariance structure of quality characteristics. Geometrically, generalised variance can be interpreted in terms of the volume of a p-dimensional parallelotope defined by p quality characteristics under study (Anderson [2]). Based on this interpretation, process variability is measured in terms of the volume of the parallelotope: the larger the volume, the larger the process variability; and the smaller the volume, the smaller the process variability.

According to the above interpretation, a generalised variance chart will not detect a change in covariance structure if there is no change in the determinant of the covariance matrix – or, equally, no change in the volume of the p-dimensional parallelotope. Therefore such a chart by itself is not sufficient for the situation. To improve the performance of this control charting procedure in what follows, a new procedure is introduced by combining a generalised variance chart and a vector variance chart. The latter chart, based on vector variance as a measure of process variability, has recently been presented in Djauhari et al. [4].

Djauhari et al. [4] illustrate the advantage of a vector variance chart relative to a generalised variance chart by comparing them in terms of the average run length based on a certain shift in covariance structure. In general, however, there are many situations where generalised variance is not able to differentiate two different covariance matrices, while vector variances can differentiate between two different covariance matrices. Conversely, there are situations where generalised variance is able to differentiate between them while vector variance is not. The following three hypothetical covariance matrices may be considered.

$$\Sigma_1 = \begin{pmatrix} 4 & 3 \\ 3 & 9 \end{pmatrix}, \ \Sigma_2 = \begin{pmatrix} 26 & 5 \\ 5 & 2 \end{pmatrix} \text{ and } \Sigma_3 = \begin{pmatrix} 10 & \sqrt{7} \\ \sqrt{7} & 1 \end{pmatrix}. \tag{1}$$

These matrices represent three differing covariance structures. The variance of the first and the second variables, and also the correlation coefficient between them represented by Σ_1, are totally different from those represented by Σ_2 and Σ_3. However, Σ_1 and Σ_2 have the same generalised variance but a different vector variance. On the other hand, Σ_1 and Σ_3 have the same vector variance but a different generalised variance.

The above properties show that if generalised variance is not able to detect abrupt shifts in a covariance structure, vector variance could possibly do so, and vice versa. It is these properties that suggest combining a generalised variance chart and a vector variance chart; one chart is used after the other. The ability of this control charting procedure to detect the shift in covariance structure will be better than one based on a generalised variance chart alone.

The combination procedure is used to control the variability of a B-complex vitamin production process in a pharmaceutical company. The result is satisfactory, and is reported here. To begin the discussion, the next section reports the experience in using a generalised stand-alone variance chart. The third section offers an analysis to see whether or not this control charting procedure is convincing. The result of this analysis motivates the use of a vector variance chart as another control charting procedure. This is presented in the fourth section. A new procedure will then be introduced to handle the limitation of the generalised variance chart. Additional remarks close this paper.

2. GENERALISED VARIANCE CHART IN CONTROLLING B-COMPLEX PRODUCTION PROCESS

In what follows, the discussion is focused on Phase I process control. For this purpose, suppose m independent samples drawn from a multivariate normal process are available. Denote S_k as the covariance matrix of sample k of p rows and p columns; $k = 1, 2, ..., m$, and $\bar{S}$ their average. A generalised variance chart consists of plotting the determinant of S_k, denoted by $|S_k|$, the lower control limit (LCL), and the upper control limit (UCL). If LCL is found to be negative, it is common to set it to 0. The control limits are used as the cut-off values to decide whether or not an out-of-control signal occurs. An out-of-control signal is declared to occur at sample k if $|S_k|$ is not in the control region. To calculate these control limits, the following formula, which ensures lack of biased is used (see Djauhari [3]).

$$\text{LCL} = |\bar{S}| \left(\frac{b_1}{b_3} - 3 \sqrt{\frac{b_2}{b_3^2 + b_4}} \right) \text{ and UCL} = |\bar{S}| \left(\frac{b_1}{b_3} + 3 \sqrt{\frac{b_2}{b_3^2 + b_4}} \right), \tag{2}$$

where: 1. $b_1 = \dfrac{1}{(n-1)^p} \prod\limits_{k=1}^{p} (n-k)$,

2. $b_2 = b_1 \left\{ \dfrac{1}{(n-1)^p} \prod\limits_{k=1}^{p} (n-k+2) - b_1 \right\}$,

3. $b_3 = \dfrac{1}{\{m(n-1)\}^p} \prod\limits_{k=1}^{p} \{m(n-1) - k + 1\}$,

4. $b_4 = b_3 \left\{ \dfrac{1}{\{m(n-1)\}^p} \prod\limits_{k=1}^{p} \{m(n-1) - k + 3\} - b_3 \right\}$.

Here $\displaystyle\prod_{k=1}^{p}(n-k) = (n-1)(n-2)\ldots(n-p)$ and the probability of false alarm, i.e., the probability that an out-of-control signal will occur even though the process is in-control, is 0.0027.

The above control charting procedure controls the production process variability of B-complex vitamin tablets in a pharmaceutical company (whose name is not revealed to preserve confidentiality). The relevant functionary at that company provided data of m = 15 independent samples, where the number of observations in each sample is n = 12 and the number of quality characteristics of p = 2 – i.e., the 'mass' and 'hardness' of the tablet. From this data, the covariance matrix and the generalised variance of each sample have been established. The results are presented in Table 1 and Table 2.

Sample k	S_k		Sample k	S_k	
1	0.24692	0.09532	9	0.28159	0.03339
	0.09532	0.40818		0.03339	0.58811
2	0.35937	-0.03545	10	0.08235	-0.02020
	-0.03545	0.29879		-0.02020	0.14636
3	0.08390	-0.01620	11	0.04546	0.04427
	-0.01620	0.19114		0.04427	0.34606
4	0.52692	0.04270	12	0.17106	-0.04770
	0.04270	0.49515		-0.04770	0.16568
5	1.22232	-0.39477	13	0.27015	0.20255
	-0.39477	0.27091		0.20255	0.24205
6	0.53114	-0.09150	14	0.38228	0.10867
	-0.09150	0.26788		0.10867	0.46811
7	0.81990	0.32914	15	0.78926	0.41982
	0.32914	0.35174		0.41982	0.50424
8	0.02674	-0.03480			
	-0.03480	0.43659			

Table 1: Covariance matrices

Sample k	$|S_k|$	Sample k	$|S_k|$
1	0.09170	9	0.16449
2	0.10612	10	0.01165
3	0.01578	11	0.01377
4	0.25910	12	0.02606
5	0.17526	13	0.02436
6	0.13391	14	0.16714
7	0.18006	15	0.22173
8	0.01047		

Table 2: Generalised variances

To construct the generalised variance chart, where the control limits defined in (2) were $\bar{S}$, calculate first its determinant $|\bar{S}|$ and the constants b_1 , b_2 , b_3 , and b_4 . From Table

1 obtain $|\bar{S}|$ = 0.13267. Furthermore, m = 15, n = 12 and p = 2, and from equation (2) b_1 = 0.90909, b_2 = 0.34561, b_3 = 0.99394, and b_4 = 0.02417. Hence, LCL = – 0.11123 which is set to 0 and UCL = 0.35392. Figure 1 shows the generalised variance chart for data in Table 2. The horizontal axis gives the sample number, and the vertical axis the generalised variance. The figure shows that all sample generalised variances are in the control region. This is an indication that, according to the generalised variance chart, no out-of-control signal occurs.
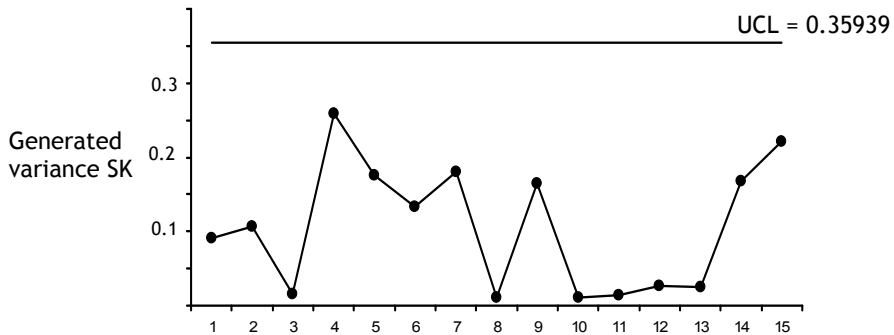

Figure 1: Generalised variance chart

3. LIMITATION OF GENERALISED VARIANCE CHART

According to the generalised variance chart in Figure 1, there is no identifiable cause of the process variability. Is this really so? This question is commonly asked by quality professionals when they are considering a control chart. To answer this question, as generalised variance is the product of all eigenvalues of the covariance matrix, which is assumed positive definite, it is recommended to conduct a comparison study of the fifteen sets, each of which consists of two eigenvalues.

Considering the covariance matrices in Table 1 anew, denote λ_{kj} the j-th eigenvalue of the k-th covariance matrix S_k; k = 1, 2, ……, 15 and j = 1 and 2. As p = 2, we call the first and second eigenvalues for the larger and smaller eigenvalues respectively. For each covariance matrix presented in Table 1, the eigenvalues are summarised in Table 3.

	Eigenvalues			Eigenvalues	
Sample	First	Second	Sample	First	Second
1	0.45240	0.20270	9	0.59170	0.27799
2	0.37571	0.28245	10	0.15220	0.07651
3	0.19352	0.08155	11	0.35245	0.03907
4	0.55659	0.46548	12	0.21617	0.12057
5	1.36482	0.12841	13	0.45913	0.05307
6	0.55982	0.23919	14	0.54203	0.30835
7	0.98971	0.18193	15	1.09009	0.20340
8	0.43950	0.02381			

Table 3: Eigenvalues of covariance matrices

Graphically, the columns 'First Eigenvalues' and 'Second Eigenvalues' in this table can be represented in Figure 2 in the form of a combined run chart. The solid line is the run chart for λ_{k1} and the dashed line is for λ_{k2}.
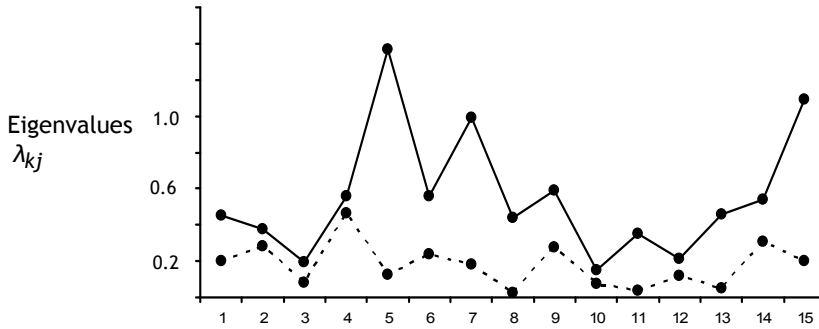
Eigenvalues λ_{kj}

Figure 2: Run chart of the k two eigenvalues

The figure is more informative than Table 3. The figure of the run chart of λ_{k1} shows a dispersed situation. It is more dispersed than that of λ_{k2}. This run chart is a powerful tool for seeing a strong indication that the covariance structure has actually shifted during the process. Considering the run of λ_{k1}, the value of λ_{51} (the fifth point from the left at the solid line) at a glance is relatively large compared with the other points. It is natural, then, to ask whether or not the covariance structure has actually shifted at sample 5 – although, according to the generalised variance chart in Figure 1, it is not an unusual point. This will be explored in greater depth in the next section by using vector variance as a further measure of process variability.

4. VECTOR VARIANCE CHART

A vector variance chart (see Djauhari et al. [4]) is another Shewhart-type chart that may be used to detect an abrupt change in process variability. Vector variance and generalised variance are two different measures of process variability. If generalised variance is the determinant of the covariance matrix, vector variance is the trace (usually denoted by 'Tr') of the squared covariance matrix – i.e., the sum of all diagonal elements of that matrix. To calculate the vector variance of sample k, $Tr\left(S_k^2\right)$, one need not find the matrix S_k^2. In practice, it is calculated as the sum of squares of all elements of S_k.

A vector variance chart plots the value of $Tr\left(S_k^2\right)$ and the control limits LCL* and UCL*. These control limits (see Djauhari et al. [4]) are given by:

$$\text{LCL*} = \hat{\theta} - 3\frac{\hat{\eta}}{\sqrt{n-1}} \text{ and UCL*} = \hat{\theta} + 3\frac{\hat{\eta}}{\sqrt{n-1}}, \tag{3}$$

where LCL* is set to 0 if it is negative, and the probability of a false alarm is 0.0027, and

1. $\hat{\theta} = \dfrac{n+1}{n-1}\left\{1+\dfrac{2}{m(n-1)}\right\}^{-1} Tr\left(\bar{S}^2\right)$, and

2. $\hat{\eta}^2 = \dfrac{8n}{n-1}\left\{1+\dfrac{12}{m(n-1)}+\dfrac{12}{\{m(n-1)\}^2}\right\}^{-1} Tr\left(\bar{S}^4\right)$

are estimates of the mean and variance of $Tr\left(S_k^2\right)$ based on m independent samples respectively. Furthermore, to calculate the statistics $Tr\left(\bar{S}^2\right)$ and $Tr\left(\bar{S}^4\right)$ one only needs to find the matrix $\bar{S}^2$. The sum of all diagonal elements of $\bar{S}^2$ is $Tr\left(\bar{S}^2\right)$, and the sum of the squares of all its elements is $Tr\left(\bar{S}^4\right)$. Based on the vector variance chart defined in (3), it is provisionally declared that an out-of-control signal at sample k occurs if $Tr\left(S_k^2\right)$ is less than LCL* or greater than UCL*.

To construct the vector variance chart for data in Table 1, first calculate the vector variance of each sample. The results are presented in Table 4.

Sample k	$Tr\left(S_k^2\right)$	Sample k	$Tr\left(S_k^2\right)$
1	0.24575	9	0.42739
2	0.22094	10	0.02902
3	0.04410	11	0.12574
4	0.52646	12	0.06127
5	1.87921	13	0.21362
6	0.37061	14	0.38888
7	1.01263	15	1.22968
8	0.19375		

Table 4: Vector variances

Then calculate the control limits. To do this, from Table 1 consecutively calculate $\bar{S}$, its square $\bar{S}^2$, the sum of all diagonal elements of $\bar{S}^2$, and the sum of squares of all elements of $\bar{S}^2$, obtaining $Tr\left(\bar{S}^2\right)$ = 0.27444 and $Tr\left(\bar{S}^4\right)$ = 0.33666. Therefore, $\hat{\theta}$ = 0.32045 and $\hat{\eta}$ = 1.65464 and thus, from equation (3), LCL* = – 1.17623 which is set to 0 and UCL* = 1.81712. Based on these results, the vector variance chart for data is shown in Figure 3. The control chart shows that the fifth sample point is beyond the control region. Its vector variance is greater than the upper control limit UCL*. Thus, according to the vector variance chart, an out-of-control signal occurs at sample 5, as indicated by the run chart of the first eigenvalue in Figure 2.
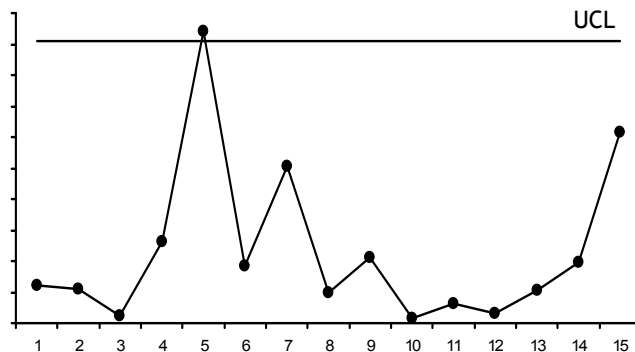

Figure 3: Vector variance chart

213

5. PROPOSED PROCEDURE

Section 2 and Section 4 show that the generalised variance chart does not give any out-of-control signal during the process, while the vector variance chart does. In practice, it is also possible to come across a situation where the generalised variance chart signals an out-of-control state but the vector variance chart does not. These properties show that if one chart is not able to detect an abrupt shift in covariance structure, the other chart could possibly do so. Due to these properties, the following procedure is proposed where both charts are used in turn. The procedure consists of two steps:

1. Select one of the two charts – a generalised variance chart or a vector variance chart. Suppose one selects a generalised variance chart. If this chart does not give any out-of-control signal, then go to the second step. Otherwise, an out-of-control signal occurs.
2. Use the alternative chart. In this case use a vector variance chart. If this chart does not signal an out-of-control state, then one may more confidently declare that an assignable cause has not occurred. Otherwise, an out-of-control signal occurs.

The procedure is presented in Figure 4 in the form of a flowchart, where GV and VV stand for generalised variance and vector variance respectively.
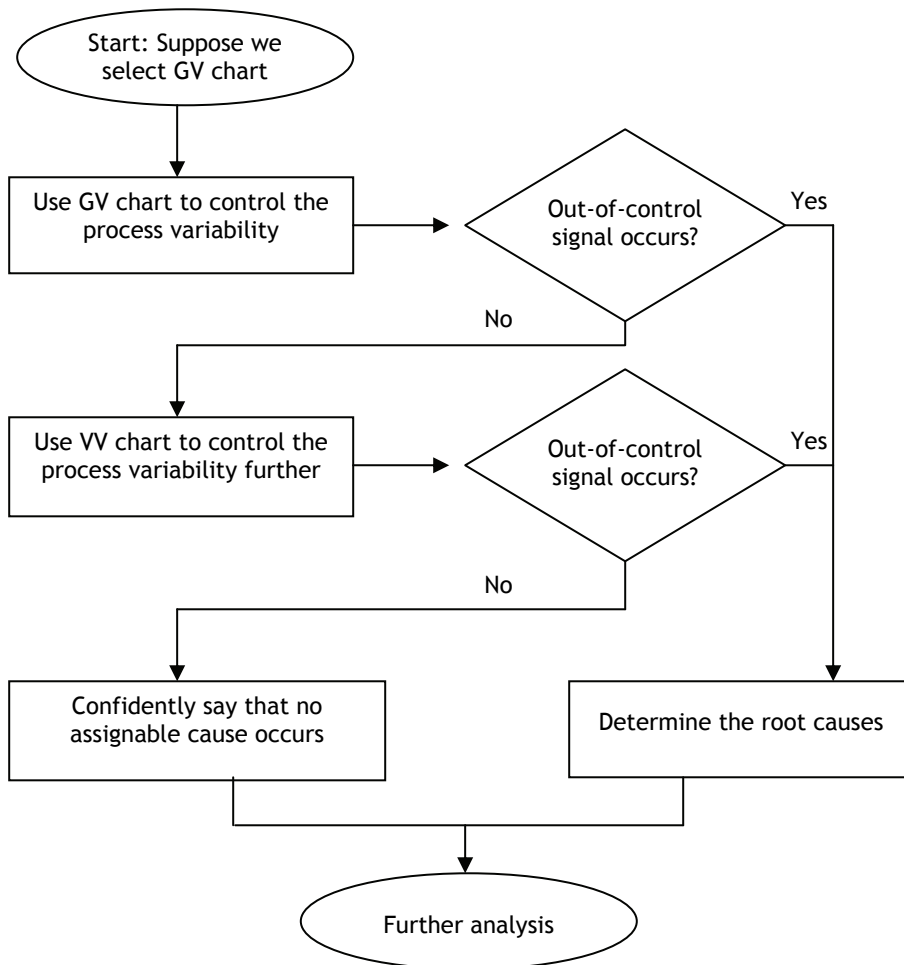

Figure 4: Flowchart of the proposed procedure

214

6. ADDITIONAL REMARKS

When the covariance structure shifts abruptly, the ability of the proposed procedure to detect that shift will be better than the generalised variance chart alone. Furthermore, considering equations (2) and (3), this procedure appears complicated. But in fact it is easy to implement, even by using very familiar software for non-statisticians such as Microsoft Excel. In this paper, all calculations and drawings were done using this software.

The use of the proposed procedure in controlling B-complex vitamin production process variability shows the situation where generalised variance chart does not signal any shift of covariance structure, while the vector variance chart does. In practice, one may find examples where a generalised variance chart signals an out-of-control state but a vector variance chart does not. Thus this procedure will decrease the probability that an out-of-control signal does not occur even though the process variability has shifted. In other words, we can be more confident that there is no assignable cause in the process if both charts do not signal any of out-of-control state. Although the experience reported in the previous sections is in a pharmaceutical company with $p = 2$, the procedure can be used in any manufacturing industry with larger p.

This paper ends with an important warning: the procedure proposed above is not free of limitations. It might happen that an out-of-control signal does not occur even though the covariance structure has actually shifted. As a hypothetical example of this situation, we compare Σ_1 in (1) with the following covariance matrix:

$$\Sigma_4 = \begin{pmatrix} 10 & \sqrt{3} \\ \sqrt{3} & 3 \end{pmatrix}.$$

The two matrices Σ_1 and Σ_4, which represent two different covariance structures, have the same generalised variance and the same vector variance. The problem of finding a procedure that is able to eliminate the error of not signaling an out-of-control state when the process variability has abruptly shifted is a research topic for the future.

7. ACKNOWLEDGEMENT

8. REFERENCES

[1] **Alt, F.B.,** & **Smith, N.D.** 1988. Multivariate process control. *Handbook of statistics*, P. R. Krishnaiah and C. R. Rao (eds), Elsevier Science Publishers, 7, pp. 333-351.
[2] **Anderson, T.W.** 2003. *An introduction to multivariate statistical analysis*, 3rd edition, John Wiley and Sons.
[3] **Djauhari, M.A.** 2005. Improved monitoring of multivariate process variability. *Journal of Quality Technology*, 37, pp. 32-39.
[4] **Djauhari, M.A., Mashuri, M.** & **Herwindiati, D.E.** 2008. A multivariate process variability. *Communication in Statistics – Theory and Methods*, 37, pp. 1742-1754.
[5] **Mason, R.L., Chou, Y.M.** & **Young, C.J.** 2009. Monitoring variation in a multivariate process when the dimension is large relative to the sample size. *Communication in Statistics – Theory and Methods*, 38, pp. 939-951.
[6] **Rooney, J.J., Kubiah, T.M., Westcott, R., Reid, R.D., Wagoner, K., Pylipow, P.E.** & **Plsek, P.** 2009. Building from the basics. *Quality Progress*, 1, pp. 18-29.