

A Resetting Scheme for Process Parameters using the Mahalanobis-Taguchi System

Changsoon Park¹

¹Department of Statistics, ChungAng University

(Received May 10, 2012; Revised June 14, 2012; Accepted July 19, 2012)

Abstract

Mahalanobis-Taguchi system(MTS) is a statistical tool for classifying the normal group and abnormal group in multivariate data structures. In addition to the classification itself, the MTS uses a method for selecting variables useful for the classification. This method can be used efficiently especially when the abnormal group data are scattered without a specific directionality. When the feedback adjustment procedure through the measurements of the process output for controlling process input variables is not practically possible, the reset procedure can be an alternative one. This article proposes a reset procedure using the MTS. Moreover, a method for identifying input variables to reset is also proposed by the use of the contribution. The identification of the root-cause parameters using the existing dimension-reduced contribution tends to be difficult due to the variety of correlation relationships of multivariate data structures. However, it became possible to provide an improved decision when used together with the location-centered contribution and the individual-parameter contribution.

Keywords: Classification analysis, reset procedure, contribution, multivariate data, normal group, threshold.

1. 서론

다변량 데이터에서 두 점간의 거리를 나타내는 도구로는 마할라노비스(Mahalanobis) 거리가 있다. P -변량 데이터의 평균과 공분산을 각각 $\boldsymbol{\mu}$ 와 $\boldsymbol{\Sigma}$ 라 할 때, 중심 $\boldsymbol{\mu}$ 에서 한 점 $\mathbf{x} = (x_1, x_2, \dots, x_p)$ 까지의 마할라노비스 거리는 다음과 같이 정의된다.

$$d = (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}). \quad (1.1)$$

식 (1.1)에서 마할라노비스 거리 d 는 p 변량 정규분포 $MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 의 결합확률밀도 함수에서 나타나는 표현으로서 이는 이 점이 발생할 수 있는 우도(likelihood)와 밀접한 관계를 가짐을 알 수 있다. 이 마할라노비스 거리는 분류분석에서 각 점을 여러 집단 중 하나에 분류할 때, 집단중심으로부터의 거리계산을 통해 가장 가까운 집단으로 분류하는데 사용되고 있다.

This research was supported by the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (2012-003545).

¹Professor, Department of Applied Statistics, Chung-Ang University, 221 Heukseok-Dong, Dongjak-Gu, Seoul 156-756, Korea. E-mail: cspark@cau.ac.kr

다구찌(Taguchi)는 마할라노비스 거리를 분류분석에 사용할 때 일반적인 가정과는 좀 다른 환경을 고려하였다. 먼저 전체 데이터를 두 개의 그룹으로 분류하는데, 한 그룹은 어떤 공통된 특성을 가지는 그룹으로 정의하고 이를 정상그룹(normal group)이라 하고, 다른 그룹은 아무런 공통된 특징 없이 정상그룹에 속하지 않는 데이터들의 집단으로 정의하고 이를 비정상그룹(abnormal group)이라 하였다. 즉, 다구찌는 이와 같이 데이터를 정상그룹과 비정상그룹으로 분류할 때 마할라노비스 거리를 사용하는 방법을 제안하였고 이를 마할라노비스-다구찌 시스템(Mahalanobis-Taguchi System; MTS)이라 명명하였다 (Taguchi와 Jugulum, 2002; Taguchi 등, 2005). 마할라노비스-다구찌 시스템의 절차는 다음과 같다.

정상 그룹과 비정상 그룹의 데이터를 각각 \mathbf{X}^N 과 \mathbf{X}^A 라 하면

$$\mathbf{X}^N = \begin{pmatrix} x_{11}^N & x_{12}^N & \cdots & x_{1p}^N \\ x_{21}^N & x_{22}^N & \cdots & x_{2p}^N \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_N1}^N & x_{n_N2}^N & \cdots & x_{n_Np}^N \end{pmatrix}_{n_N \times p}, \quad \mathbf{X}^A = \begin{pmatrix} x_{11}^A & x_{12}^A & \cdots & x_{1p}^A \\ x_{21}^A & x_{22}^A & \cdots & x_{2p}^A \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_A1}^A & x_{n_A2}^A & \cdots & x_{n_Ap}^A \end{pmatrix}_{n_A \times p}$$

로 표시한다. 이 때 정상그룹과 비정상 그룹의 데이터 수는 각각 n_N 과 n_A 가 된다. 정상 그룹과 비정상 그룹에서 각각의 데이터는

$$\mathbf{X}_i^N = (x_{i1}^N, x_{i2}^N, \dots, x_{ip}^N), \quad i = 1, 2, \dots, n_N,$$

$$\mathbf{X}_j^A = (x_{j1}^A, x_{j2}^A, \dots, x_{jp}^A), \quad j = 1, 2, \dots, n_A$$

로 표현한다. 전체 데이터를 \mathbf{X} 라 하면

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^N \\ \mathbf{X}^A \end{pmatrix}$$

가 된다. 이 때 전체 데이터의 수는 $n = n_N + n_A$ 이다. 정상 그룹의 평균벡터를 $\boldsymbol{\mu}^N = (\mu_1^N, \mu_2^N, \dots, \mu_p^N)$, 공분산 행렬을 $\boldsymbol{\Sigma}^N$ 이라 하면 마할라노비스-다구찌 거리(Mahalanobis-Taguchi distance)는 정상 그룹과 비정상그룹에 대해 다음과 같이 정의한다.

$$d_i^N = (\mathbf{X}_i^N - \boldsymbol{\mu}^N) (\boldsymbol{\Sigma}^N)^{-1} (\mathbf{X}_i^N - \boldsymbol{\mu}^N)', \quad i = 1, 2, \dots, n_N,$$

$$d_j^A = (\mathbf{X}_j^A - \boldsymbol{\mu}^N) (\boldsymbol{\Sigma}^N)^{-1} (\mathbf{X}_j^A - \boldsymbol{\mu}^N)', \quad j = 1, 2, \dots, n_A. \quad (1.2)$$

식 (1.2)에서 마할라노비스-다구찌 거리를 계산할 때 정상 그룹과 비정상 그룹에 관계없이 정상 그룹의 중심 ($\boldsymbol{\mu}^N$)으로부터의 마할라노비스 거리를 계산함을 주의할 필요가 있다. 이 때 그룹의 평균과 공분산 행렬을 계산하는 정상그룹에 속한 데이터들의 집합을 마할라노비스 공간(Mahalanobis space)이라고 한다. 마할라노비스 공간은 정상그룹 데이터 전체가 될 수도 있고 또는 그 일부만 될 수도 있으나 판단은 환경에 따라 다르다. 마할라노비스-다구찌 시스템에 의한 분류방법은, 어떤 경계 값(threshold) T 에 대하여 마할라노비스 거리 d 가 $d \geq T$ 이면 비정상 그룹으로 분류하고, 그렇지 않으면 정상 그룹으로 분류한다.

마할라노비스-다구찌 시스템에서는 분류에 도움이 되는 변수만을 선택하기 위해 수준수가 2(0, 1로 표시)인 직교배열표(orthogonal array)와 망대특성(larger-the-better type) 신호대잡음비(signal-to-noise

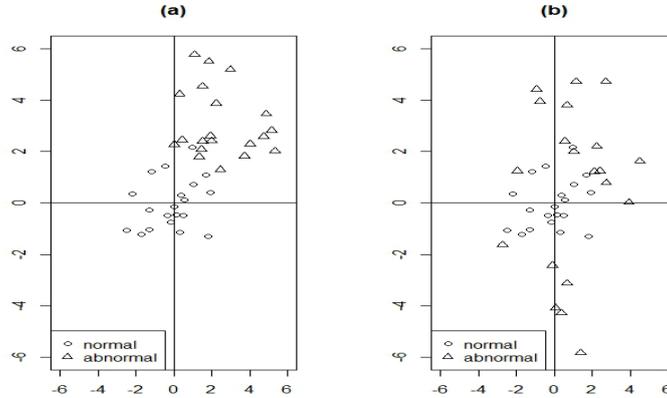


Figure 2.1. The distribution of the abnormal group data according to the direction-variant (a) and the direction-invariant (b)

ratio)를 사용하였다. 이 방법은 $L_{k+1}(2^k)$ 직교배열표의 열(column)에 변수를 배치하고 배열표의 값이 0이면 변수를 포함하지 않고 1이면 포함한다. 그런 다음 $(k+1)$ 개의 실험에서 각각 비정상 그룹의 데이터에 대한 마할라노비스 거리를 계산하고 이에 대한 망대특성 신호대 잡음비를 통해 각 변수의 gain(변수를 포함할 때와 하지 않을 때의 신호대잡음비 평균의 차)이 양수인 변수를 분류에 도움이 되는 변수로 판단하는 것이다. 이와 같은 변수선택법을 사용하면 변수의 수를 줄이면서도 효과적인 분류를 할 수 있다.

마할라노비스-다구찌 시스템은 Kanetaka (1988)의 환자 분류에서 시작되어 오늘에 이르고 있으나 이에 대한 연구는 큰 주의를 끌지 못하고 또한 많은 연구도 진행되지 않았다. Woodall 등 (2003)은 이 방법에 대한 비평을 통해 마할라노비스-다구찌 시스템이 분류분야에서 큰 공헌을 하지 못함을 표현하기도 하였다. 그러나 Taguchi와 Jugulum (2002)은 마할라노비스-다구찌 시스템을 기존의 분류방법과 비교할 때 신호대잡음비를 사용하여 변수를 선택할 수 있고, 제곱손실을 고려하여 경계값을 설정할 수 있다는 차이점을 들며 이 방법이 분류분야에 기여하고 있음을 설명하였다. 또한 공정관리에서 이상원인의 탐지에 마할라노비스-다구찌 시스템을 이용할 것을 추천하고 있다.

2. 마할라노비스-다구찌 시스템의 특성

분류분석에는 여러 종류의 분류방법이 존재하며 이 분야에서 효율적으로 잘 사용되어 오고 있다. 그룹의 수가 k 개일 때 각 그룹이 나름의 군집을 형성하고 있으면 기존의 분류방법으로 충분한 분류효과를 얻을 수 있다. 특히 그룹의 수가 $k = 2$ 일 때는 분류가 효과적이면서 용이하다. 하지만 $k = 2$ 일 때 한 그룹은 나름의 군집을 형성하지만 다른 한 그룹이 군집을 형성하지 않고 단순히 첫 번째 그룹에 속하지 않는 것으로 구분될 때에는 기존의 분류방법이 그 역할을 못하는 경우가 있다. 이에 대한 예는 Figure 2.1을 통해 설명한다. Figure 2.1(a)는 두 그룹이 각각 군집을 형성하는 경우이다. 이 경우에 판별분석을 사용하면 대부분의 점들이 정분류(correct classification)가 됨을 알 수 있다. 그러나 Figure 2.1(b)는 한 그룹은 군집을 형성하는데 다른 그룹은 군집을 형성하지 않고 단순히 첫째 그룹에서 멀리 떨어져 있는 점들로 형성되어 있는 경우이다. 즉, 데이터들이 마할라노비스 공간으로부터 일정한 방향성이 없이 흩어져 있는 경우이다. 이러한 경우에 판별분석을 통한 분류는 오분류의 가능성이 너무 커서 분류방법으로 적절치 못함을 알 수 있다. 이러한 경우에 마할라노비스-다구찌 시스템을 사용하면 아주

잘 분류를 할 수 있게 된다. 이와 같이 마할라노비스-다구찌 시스템은 다른 그룹이 방향성이 없이 흩어져 있을(directionally invariant) 때에 효율적으로 사용할 수 있는 분류방법이다.

이를 다른 관점에서 보면 비정상 그룹은 마할라노비스 공간의 중심으로부터 일정 거리 이상 떨어져 있으나 각 변수의 값이 그 변수의 중심 값보다 한 방향이 아닌 클 수도 작을 수도 있어 하나의 군집(cluster)을 형성하지 못하는 경우에 효율적이다. 반대로 말하면 정상그룹과 비정상 그룹이 각각 하나의 군집을 형성하는 경우에는 기존의 다른 분류방법과 비교할 때 특별히 효율적이지는 않다.

마할라노비스-다구찌 시스템의 또 다른 특징은 마할라노비스 공간이 정의되면 비정상 그룹에 대한 데이터가 없어도 분류방법을 정의할 수 있다는 점이다. 즉, 마할라노비스 공간으로부터 마할라노비스 거리를 정의하고 적절한 경계값을 설정하면 분류방법이 정의될 수 있으며, 이 때 경계값은 마할라노비스 공간에서 정의된 마할라노비스-다구찌 거리의 분포를 이용하여 설정할 수 있다. 따라서 마할라노비스-다구찌 시스템은 기존의 분류문제에서 이러한 효율성을 나타낼 수 있는 경우에 적용해볼 가치가 있다. 그 한가지 예로 통계적 공정관리 분야를 들 수 있다.

3. 마할라노비스-다구찌 시스템을 이용한 공정모수의 재조정 절차

다단계 제조과정(multistage manufacturing process; MMP)은 여러 단위공정(unit process)이 직렬로 연결되어 하나의 전체 공정(whole process)을 형성한다. 각 단위공정에는 여러 입력모수(input parameter)들이 있으며, 공정 제조환경을 최적으로 하기 위해 이 입력모수들이 특정한 값(또는 수준)으로 설정되어 있다. 이 값들이 관리자의 의도대로 설정이 되면 관리요인(control factor; \mathbf{z})이라 하고 그렇지 않고 그 수준이나 설정 값이 조정하기 힘들거나 비용이 과도하게 들면 잡음요인(noise factor; \mathbf{x})이라 한다. 잡음요인은 설정된 값을 중심으로 변동을 가지고 있어 이는 하나의 확률변수로 간주하여 그 통계적 특성(평균과 분산 등)으로 알려져 있거나 지정되지만 실제 어떤 값을 가지는지는 알 수 없다. 또한 여러 개의 잡음요인들은 서로 상관관계를 가지고 변하여 다변량분석을 통한 변화의 탐지가 필요하다.

한 단위 공정의 입력 잡음요인으로 p 개의 서로 상관된 변수를 고려하자. 이 변수들은 공정의 출력값 $y = f(\mathbf{x}; \mathbf{z})$ 에 영향을 주고 있으나 그 영향을 나타내는 식은 알려있지 않은 경우가 대부분이다. 이런 경우에는 관리된 입력요인 하에서 이에 따른 출력값을 관측하여 통계적 방법(회귀분석이나 실험계획)으로 관리할 수 밖에 없다. 그러나 다단계 공정에서는 각 단위공정의 출력값을 측정하지 못하는 경우가 많다. 또는 측정하더라도 오랜 시간이 경과하여 피드백 수정(feed-back adjustment)이 비효율적이 된다. 따라서 이와 같은 경우에는 피드포워드 수정(feed-forward adjustment)같이 공정의 입력단계에서 입력요인을 관리하는 방법을 생각할 수 있다. 하지만 피드포워드 수정이 가능하려면 입력요인의 값에 따른 출력값을 추정할 수 있어야 하는데 다단계 공정에서는 이마저도 불가능한 경우가 대부분이다. 이와 같은 공정여건에서 발생하는 문제를 해결하기 위해 생각할 수 있는 방법은 공정 요인을 재조정(reset)하는 것이다. 공정 수정(process adjustment)은 입력요인의 값을 일정 양만큼 작게 또는 크게 하여 출력값을 원하는 수준으로 얻기 위한 통계적 추정 방법에 해당하지만, 재조정(reset)은 공정 모수값을 공정의 시작 때처럼 새롭게 설정을 하는 작업이다. 공정 재조정 문제는 관리도에서 이상원인 탐색의 문제점을 개선하기 위해 Park (2012)이 제안하였으며 이러한 재조정 절차는 공정수정 문제에도 적용될 수 있다 (Park, 2013).

마할라노비스-다구찌 시스템을 이용한 공정 재조정 절차는 다음과 같이 설명된다. 공정의 준비작업이 시작되고 일정시간이 지난 후 공정이 안정되었다고 판단되면 생산을 시작한다. 초기에는 일정 시간 간격으로 크기 n (≥ 2)인 표본 데이터를 일정 수 (m)만큼 관측하여 이를 정상그룹의 데이터로 간주한다. 이 기간은 관리도의 적용에서 제 1국면(phase I)에 해당한다. 이 기간에 관측된 i 번째 표본데이터

를 \mathbf{X}_{0i} 라 하면 다음과 같이 표현된다.

$$\mathbf{X}_{0i} = \begin{pmatrix} x_{i11} & x_{i12} & \cdots & x_{i1p} \\ x_{i21} & x_{i22} & \cdots & x_{i2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{in1} & x_{in2} & \cdots & x_{inp} \end{pmatrix}_{n \times p}.$$

관측된 i 번째 표본데이터에 대해 각 변수들의 평균과 분산-공분산은

$$\begin{aligned} \bar{x}_{ik} &= \frac{1}{n} \sum_{j=1}^n x_{ijk}, \quad k = 1, 2, \dots, p, \\ s_{ik}^2 &= \frac{1}{n-1} \sum_{j=1}^n (x_{ijk} - \bar{x}_{ik})^2, \quad k = 1, 2, \dots, p, \\ s_{ihl} &= \frac{1}{n-1} \sum_{j=1}^n (x_{ijh} - \bar{x}_{ih})(x_{ijl} - \bar{x}_{il}), \quad h \neq l \end{aligned}$$

으로 표기되어 평균벡터와 분산-공분산 행렬은

$$\begin{aligned} \bar{\mathbf{X}}_{0i} &= (\bar{x}_{i1}, \bar{x}_{i2}, \dots, \bar{x}_{ip}), \\ \mathbf{S}_i &= \begin{pmatrix} S_{i1}^2 & S_{i12} & \cdots & S_{i1p} \\ S_{i12} & S_{i2}^2 & \cdots & S_{i2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{i1p} & S_{i2p} & \cdots & S_{ip}^2 \end{pmatrix} \end{aligned}$$

로 표현한다. 관측된 모든 m 개의 표본에 대해 총평균과 총 분산-공분산은

$$\bar{\mathbf{X}}_0 = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p), \quad (3.1)$$

$$\bar{\mathbf{S}} = \begin{pmatrix} \bar{S}_1^2 & \bar{S}_{12} & \cdots & \bar{S}_{1p} \\ \bar{S}_{12} & \bar{S}_2^2 & \cdots & \bar{S}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{S}_{1p} & \bar{S}_{2p} & \cdots & \bar{S}_p^2 \end{pmatrix} \quad (3.2)$$

로 표기한다.

정상그룹으로부터 마할라노비스 거리를 계산하면

$$d(\mathbf{X}_{0i}) = (\bar{\mathbf{X}}_{0i} - \bar{\mathbf{X}}_0) \bar{\mathbf{S}}^{-1} (\bar{\mathbf{X}}_{0i} - \bar{\mathbf{X}}_0)', \quad i = 1, 2, \dots, m$$

이 된다. 제 1국면 이후에 관측되는 데이터 $\{\mathbf{X}_{1i}, i = 1, 2, \dots\}$ 에 대해 그 평균을 \bar{X}_{1i} 라 하면 이 데이터에 대한 마할라노비스 거리는 위식의 마할라노비스 공간의 평균벡터 식 (3.1)과 분산-공분산 행렬 식 (3.2)를 사용하여 다음과 같이 표현된다.

$$d(\mathbf{X}_{1i}) = (\bar{\mathbf{X}}_{1i} - \bar{\mathbf{X}}_0) \bar{\mathbf{S}}^{-1} (\bar{\mathbf{X}}_{1i} - \bar{\mathbf{X}}_0)', \quad i = 1, 2, \dots$$

마할라노비스-다구찌 시스템을 이용한 공정모수의 재조정 절차는

$$d(\mathbf{X}_{1i}) \geq T$$

이면 공정모수를 수정한다.

4. 경계 값의 설정

재조정 절차에서 경계 값 T 를 설정하는 문제를 생각해 보자. 재조정 절차의 사용을 위해서는 T 를 결정 하여야 하는데 이는 데이터가 마할라노비스 공간의 중심에서 어느 정도 떨어져 있으면 비정상 그룹으로 간주하느냐의 문제로서 재조정 절차의 제 1종 오류를 결정하는 문제와 동일하다. 흔히 관리도에서는 제 1종 오류가 오경보에 해당되어 아주 미세한 값 (3σ 한계를 사용하는 경우 0.0027)으로 설정하는 것이 보편적 방법이다. 재조정 절차에서는 이상원인의 탐색과정이 없기 때문에 오경보와 같은 심각한 부작용대신 단지 불필요한 추가적 재조정을 유발할 가능성이 있기 때문에 이 확률을 어느 정도 크게 하면서 제 2종 오류를 줄여 좋은 출력 값을 생산하는 것이 더 바람직하다. 또한 공정 모수의 재조정절차에서는 모수의 관측 빈도가 관리도보다 품질특성치의 관측을 훨씬 자주하게 된다. 즉 재조정절차의 설계에 유용한 데이터의 수가 많다고 할 수 있다.

재조정 절차의 제 1종 오류를 α 로 하기 위해서는 정상그룹의 마할라노비스 거리의 순서통계량을 사용하여 설정할 수 있다. 먼저 정상그룹의 마할라노비스 거리의 순서통계량을

$$d_{(1)}, d_{(2)}, \dots, d_{(m)}$$

이라 하면, 가우스수 $g = [m(1 - \alpha + 1/(2m))]$ 에 대해 표본 마할라노비스 거리의 상위 α 분위수는

$$T_\alpha = d_{(g)} + m(d_{(g+1)} - d_{(g)}) \left(1 - \alpha - \frac{g - 0.5}{m}\right)$$

을 설정한다 (부록 참조). 단 이때 $\alpha > 1/m$ 이어야 한다.

마할라노비스-다구찌 시스템에서는 망대특성(larger-the-better type) 신호대잡음비(signal-to-noise ratio)와 $L_{k+1}(2^k)$ 직교배열표(orthogonal array)를 사용하여 공정모수 재조정에 도움이 되지 않는(unuseful) 모수를 선별하여 모수의 차수를 축약하는 방법을 사용하였다. 단, 이때에는 비정상 그룹의 데이터가 있어야 한다. 따라서 공정 모수의 재조정 절차에서 차수의 축약은 일정시간이 경과한 후 비정상 데이터가 누적되어 정상그룹의 데이터와 비교가 가능한 크기가 된 경우에만 가능하다.

5. 재조정 모수의 선택

위의 재조정 절차에서 재조정할 경우에는 p 개의 모수 중 설정값에서 과도하게 벗어나 마할라노비스 거리가 재조정 한계보다 크게 하는데 기여한 변수들만을 재조정할 수 있다면 모수 전체를 재조정하는 것 보다 바람직하다. 다만 p 개의 모수들이 서로 상관관계를 가지고 있기 때문에 이를 고려한 기여도(contribution)를 산정할 필요가 있다. 따라서 기여도를 산정할 때에는 개개의 모수 뿐만 아니라 두 모수에 의한 공동 기여도도 고려할 필요가 있다. 이들 모수의 기여도는 다양한 방법으로 측정할 수 있다. 그 중 Runger 등 (1996)은 p 개의 모든 모수를 사용하여 계산한 마할라노비스 거리와 특정한 하나의 모수를 제외하고 $(p - 1)$ 개의 모수를 사용하여 계산한 마할라노비스 거리와의 차로써 특정한 하나의 모수에 대한 기여도를 나타내었다. 이 논문에서는 특정한 하나의 모수를 제외하는 경우뿐만 아니라 두 개의 모수를 제외하는 경우도 추가로 고려하였다. 이는 두 모수간의 상관관계에 의한 기여도도 고려하기 위함이다. 이러한 방법은 모수의 수가 $(p - 1)$ 또는 $(p - 2)$ 로 줄어드는 경우의 마할라노비스 거리를 고려하였기 때문에 차원축소 기여도(dimension-reduced contribution; DRC)라 하고 다음과 같이 표현한다.

모든 모수를 다 사용하여 계산된 마할라노비스 거리를 d , 모든 모수들 중에서 i 번째 모수만을 제외하고 계산된 마할라노비스 거리를 $d_{-\{i\}}$, i 와 j 번째 모수를 제외하고 계산된 마할라노비스 거리를 $d_{-\{i,j\}}$ 라

하자. 하나의 모수, 즉 i 번째 모수에 의한 일모수(one parameter) 차원축소 기여도는

$$R_i^D = d - d_{-\{i\}}, \quad i = 1, 2, \dots, p.$$

두 개의 모수, 즉 i 와 j 번째 모수에 의한 이모수(two parameter) 차원축소 기여도는

$$R_{ij}^D = d - d_{-\{i,j\}}, \quad i \neq j (= 1, 2, \dots, p)$$

로 정의한다. 기여도가 크다는 것은 특정 하나 또는 두 개의 모수가 마할라노비스 거리를 크게 하는데 기여하는 상대적 크기를 의미한다. 따라서 재조정 절차에 의해 재조정 신호가 발령되면 모든 $p(p+1)/2$ 개의 기여도를 비교하여 상대적으로 큰 모수들에 대해 재조정의 필요가 있는지를 검토해야 한다. 다만 여기서 제안하기는 한꺼번에 여러 모수를 재조정하는 것보다는 기여도가 가장 큰 하나(일모수 기여도로부터) 또는 두 개(이모수 기여도로부터)의 모수만을 재조정하는 것이 바람직하다. 이와 같이 일차 재조정 후에 다음 관측값에 대한 마할라노비스 거리를 통해 재조정의 효과를 관측할 수 있다. 재조정 후의 마할라노비스 거리가 경계 값보다 작으면 재조정이 적절한 조치임을 알 수 있고 그렇지 않으면 추가적인 재조정이 반복되어야 한다. 차원축소 기여도의 계산에서는 제외되는 모수를 고려하지 않고 마할라노비스 거리를 계산하기 때문에 공분산 행렬의 차수도 줄어들며 따라서 남아있는 모수와 제외되는 모수간의 상관관계도 고려되지 못함을 알 수 있다.

이 기여도의 효용성을 알아보기 위해 다음의 예를 들어보자. 공정 모수로서 $p = 4$ 개의 잡음요인이 있을 때 서로 상관관계를 가지고 있으면서 분산이 알려져 있는 경우 상관 행렬이

$$\Sigma^N = \begin{pmatrix} 1.000 & -0.424 & 0.957 & 0.922 \\ -0.424 & 1.000 & -0.540 & -0.063 \\ 0.957 & -0.540 & 1.000 & 0.822 \\ 0.922 & -0.063 & 0.822 & 1.000 \end{pmatrix}$$

와 같다고 하자. 공정 모수의 평균, 분산 또는 상관관계의 변화를 알아보기 위해 이상상태(out of control; OC)를 나타내는 모수로서 다음을 고려한다.

$$\text{OC1} : \mu = (2, 0, 0, 0),$$

$$\text{OC2} : \mu = (2, 0, 0, 3),$$

$$\text{OC3} : \sigma_3^2 = 9.$$

관리상태일 때 평균이 $\mu = (0, 0, 0, 0)$, 공분산이 Σ^N 인 다변량 정규분포로부터 $m = 50$ 개의 데이터를 얻고 이 후의 $n_1 = 50$ 개의 데이터는 세 종류의 이상상태 환경에서 얻어졌다. 이 데이터를 통해 마할라노비스 거리를 얻은 결과는 Figure 5.1과 같다. Figure 5.1에서 보면 이상원인의 발생 후에는 마할라노비스 거리가 발생전보다 확연히 커져 재조정 신호가 신속하게 발령될 수 있음을 보이고 있다. 세 종류의 이상원인에 대한 마할라노비스 거리를 비교해 보면 평균의 변화(OC1, OC2)인 경우가 분산(OC3)의 변화보다 일관되게 커져 훨씬 용이하게 탐지될 수 있으나 분산의 변화는 상대적으로 신속한 탐지가 어려움을 나타내고 있다. 또한 OC2에서는 4번째 모수의 평균이 OC1에 비해 추가적으로 더 크게 변화하였지만 그것이 마할라노비스 거리를 오히려 더 작게하는 결과를 초래하여 일반적인 예상을 뒤엎는 결과를 보여주고 있다. 이러한 현상은 다변량 분포에서의 모수간 상관관계에 기인한 것으로 일관된 분석을 어렵게 하고 있다.

세 종류의 이상원인에 따른 차원축소 기여도는 Figure 5.2, Figure 5.3, Figure 5.4와 같다. 각 그림에서는 차원축소 기여도가 행렬로 배치되어 있다. 즉, (i, j) 번째 그림은 R_{ij}^D 이다. Figure 5.2에서는 첫 번째

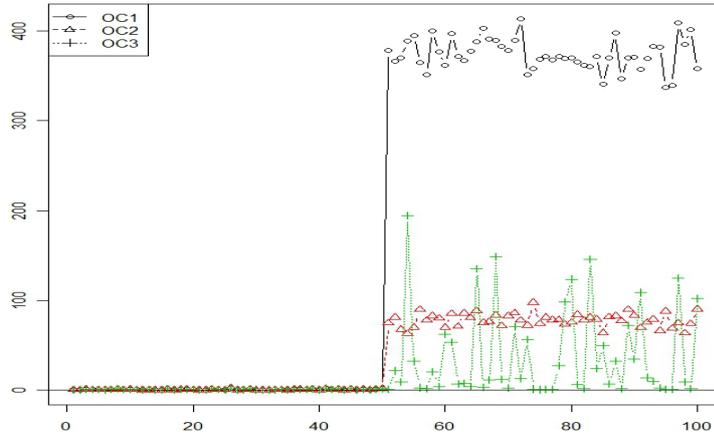


Figure 5.1. The changes in the Mahalanobis distance according to assignable causes OC1, OC2 and OC3

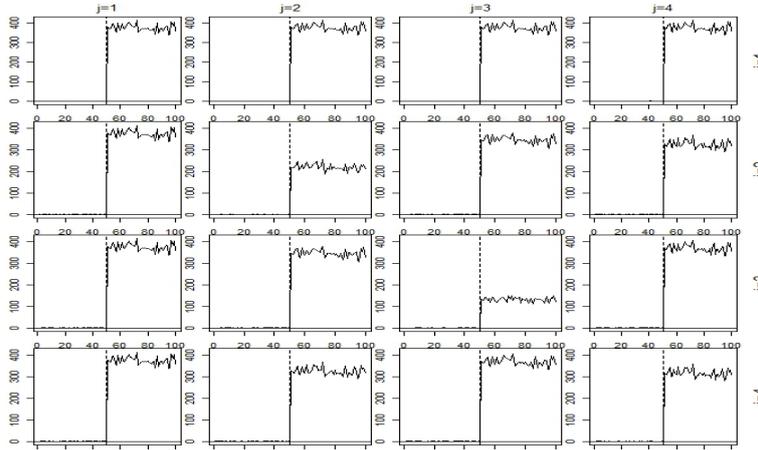


Figure 5.2. The changes in the dimension-reduction contribution according to the assignable cause OC1

평균만이 표준편차의 2배가 되었으나 기여도에서는 R_1^D 뿐만 아니라 그와 관련된 $R_{12}^D, R_{13}^D, R_{14}^D$ 가 다 같이 커지는 경향을 나타냄을 알 수 있다. 뿐만 아니라 첫 번째 모수와 직접적 관련이 없는 나머지 모든 기여도도 같이 커지는 경향이 있어 구체적으로 어느 모수에 변화가 발생하였는지를 알기 어렵다. 이와 같이 특정 모수의 변화에 직접적으로 관계없는 모수의 기여도가 커지는 현상은 이상상태 OC2, OC3의 경우에도 유사하게 나타난다 (Figure 5.3, Figure 5.4 참조). 즉, 차원축소 기여도를 사용하여서는 이상상태를 유발한 특정 모수를 지적하기에는 어려운 점이 있음을 알 수 있다.

6. 중심위치 기여도와 개별모수 기여도

앞 절에서 차원축소 기여도는 이상상태를 유발한 특정 모수를 지적하기에는 크게 효율적이지 못함을 보였다. 이와 같은 문제점은 기여도를 알아보려고 하는 모수를 완전히 제외한 다음 마할라노비스 거리를

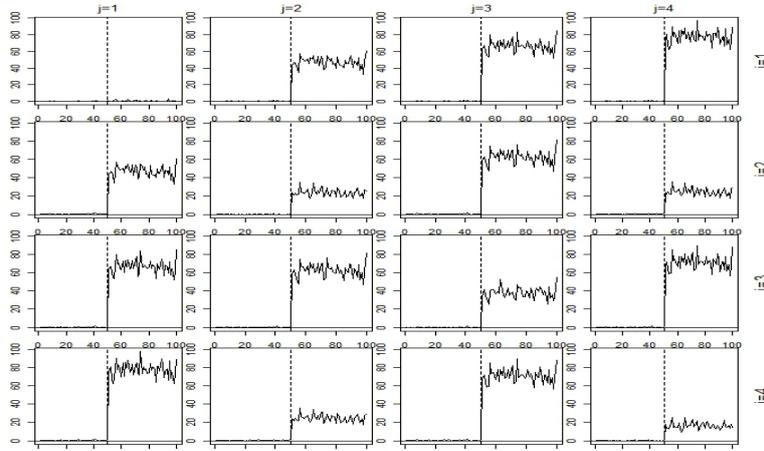


Figure 5.3. The changes in the dimension-reduction contribution according to the assignable cause OC2

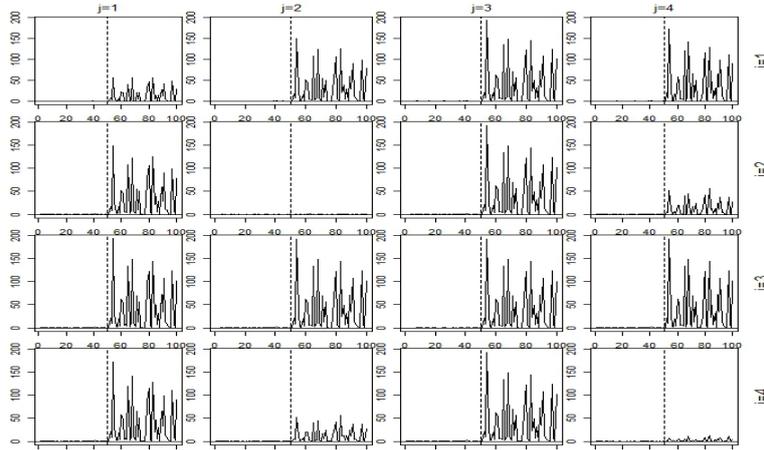


Figure 5.4. The changes in the dimension-reduction contribution according to the assignable cause OC3

계산하기 때문에 제외된 모수와 남아있는 모수 간의 상관관계가 전혀 고려되지 않기 때문에 발생한다고 볼 수 있다. 이 같은 단점을 보완하기 위해 기여도를 알아보려고 하는 모수를 완전히 제외하는 대신 그 모수를 여전히 포함하지만 그 모수의 값은 마할라노비스 공간의(즉, 정상그룹일 때) 중심값으로 지정한 다음 마할라노비스 거리를 계산하는 방법이다. 이렇게 마할라노비스 거리를 계산할 때에는 공분산 행렬을 그대로 사용하여 모든 모수간의 상관관계가 여전히 고려되는 장점이 있다. 이 기여도는 중심위치 기여도(location-centered contribution; LCC)라 하고 다음과 같이 표현한다.

모든 모수들 중에서 i 번째 모수만을 그 모수의 중심값으로 지정하고 계산된 마할라노비스 거리를 $d_{c\{i\}}$, i 와 j 번째 모수를 그 모수의 중심값으로 지정하고 계산된 마할라노비스 거리를 $d_{c\{i,j\}}$ 라 하자. 하나의 모수, 즉 i 번째 모수에 의한 일모수(one parameter) 중심위치 기여도는

$$R_i^L = d - d_{c\{i\}}, \quad i = 1, 2, \dots, p.$$

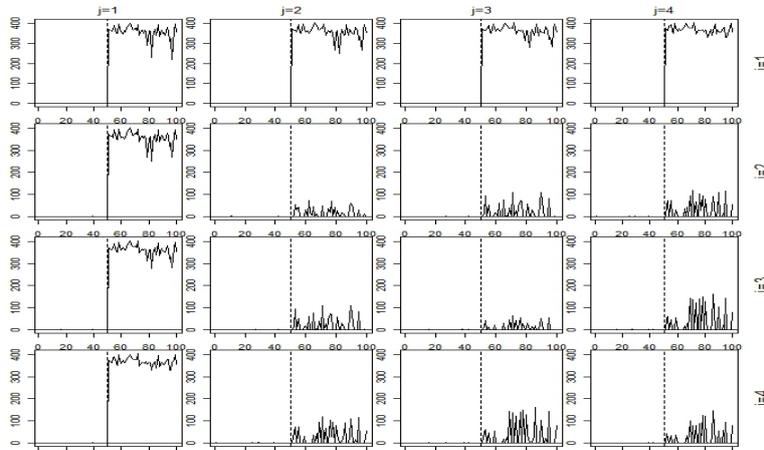


Figure 6.1. The changes in the location-centered contribution according to the assignable cause OC1

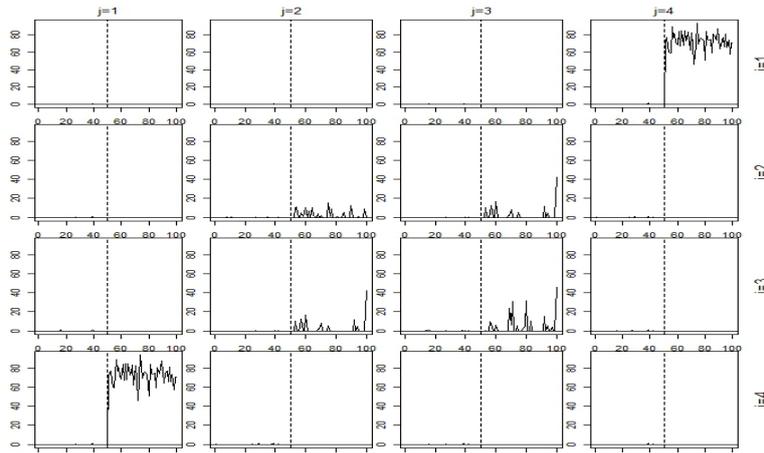


Figure 6.2. The changes in the location-centered contribution according to the assignable cause OC2

두 개의 모수, 즉 i 와 j 번째 모수에 의한 이모수(two parameter) 기여도는

$$R_{ij}^L = d - d_{c\{i,j\}}, \quad i \neq j (= 1, 2, \dots, p)$$

로 정의한다. 이 경우 마할라노비스 거리를 계산할 때 공분산 행렬은 동일한 행렬을 사용하기 때문에 남아있는 모수와 제외되는 변수간의 상관관계가 반영됨을 알 수 있다.

중심위치 기여도의 효용성을 알아보기 위해 이를 앞질의 예에 적용해 보자. 세 종류의 이상원인에 따른 개별모수 기여도는 Figure 6.1, Figure 6.2, Figure 6.3과 같다. 중심위치 기여도의 값은 음수가 될 수 있으나 이 때에는 그 모수의 기여도가 낮은 것을 의미하므로 그림에서는 기여도가 높은 경우를 위주로 판단하기 위해 음수는 모두 0으로 표현하였다. 각 그림의 (i, j) 번째 그림은 R_{ij}^L 이다. Figure 6.1은 이상원인 OC1의 경우인데 중심위치 기여도에서는 앞 질의 차원축소 기여도에서와 마찬가지로 R_{i1}^L 뿐만 아니라 그와 관련된 $R_{12}^L, R_{13}^L, R_{14}^L$ 가 다 커지는 경향을 나타냄을 알 수 있다. 하지만 첫 번째 모수

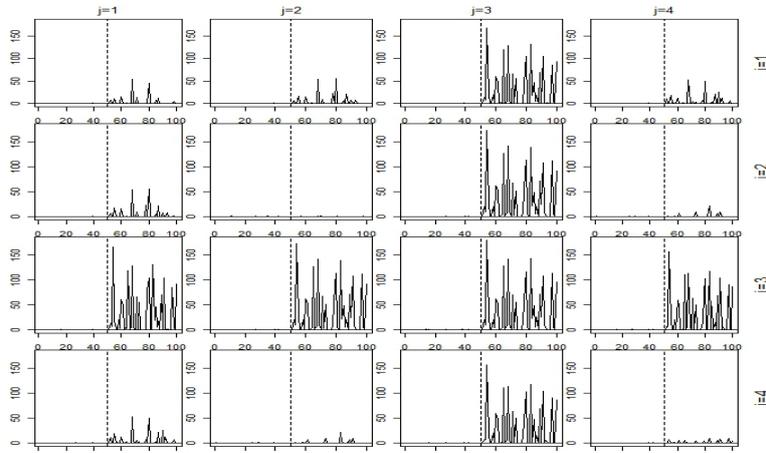


Figure 6.3. The changes in the location-centered contribution according to the assignable cause OC3

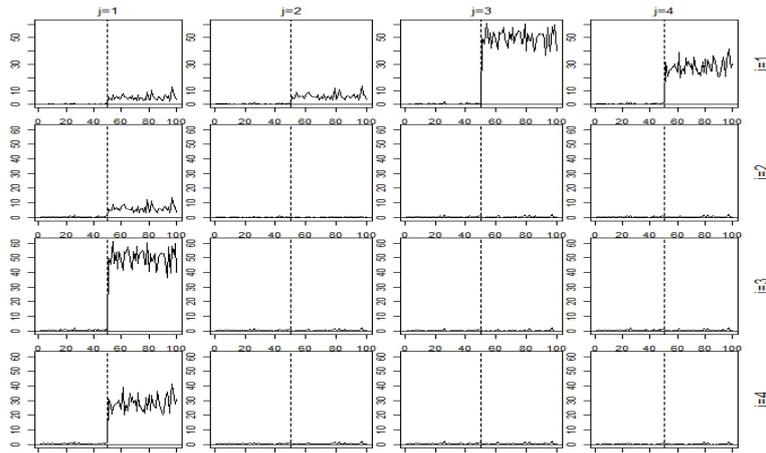


Figure 6.4. The changes in the individual-parameter contribution according to the assignable cause OC1

와 직접적 관련이 없는 나머지 중심위치 기여도는 두드러지게 커짐을 보이지 않기 때문에 구체적으로 첫 번째 모수와 관련하여 변화가 발생하였음을 알 수 있다. Figure 6.2에서는 이상원인 OC2인 경우인데 유독 (1,4)번째 그림만이 두드러지게 커짐을 보이고 있어 첫 번째 모수와 네 번째 모수가 공동으로 마할라노비스 거리를 크게 하는데 기여하고 있음을 암시하고 있다. Figure 6.3에서는 이상원인 OC3의 경우인데 $i = 3$ 인 경우와 $j = 3$ 인 경우에만 두드러지게 커짐을 보이고 있어 세 번째 모수의 기여도가 유의함을 나타내고 있다. 이와 같이 중심위치 기여도는 변화를 유발한 모수와 관계가 없는 모수에 대해서는 기여도가 커지지 않아 차원축소 기여도와 비교할 때 상대적으로 원인 모수를 탐지하는데 도움이 됨을 알 수 있다.

중심위치 기여도를 보완하는 장치로는 개별모수(individual parameter) 기여도를 고려할 수 있다. 개별모수 기여도는 다른 모수는 고려하지 않고 오직 보고자 하는 모수만을 사용하여 마할라노비스 거리를 계산하고 그 크기를 기여도로 간주하는 것이다. 하나의 모수, 즉 i 번째 모수에 의한 일모수(one param-

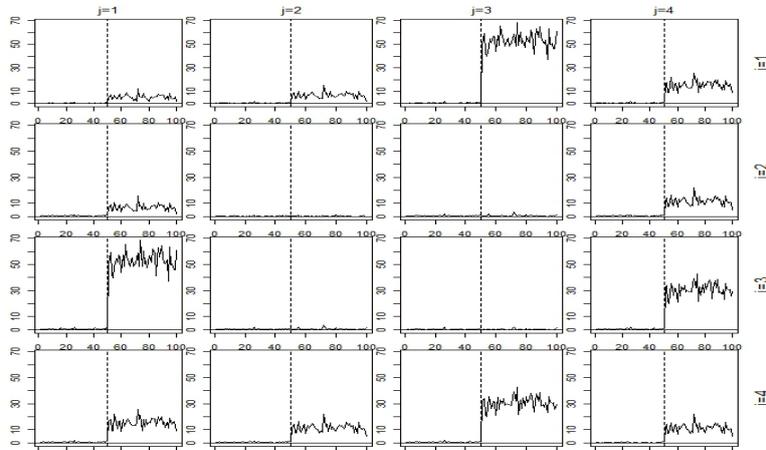


Figure 6.5. The changes in the individual-parameter contribution according to the assignable cause OC2

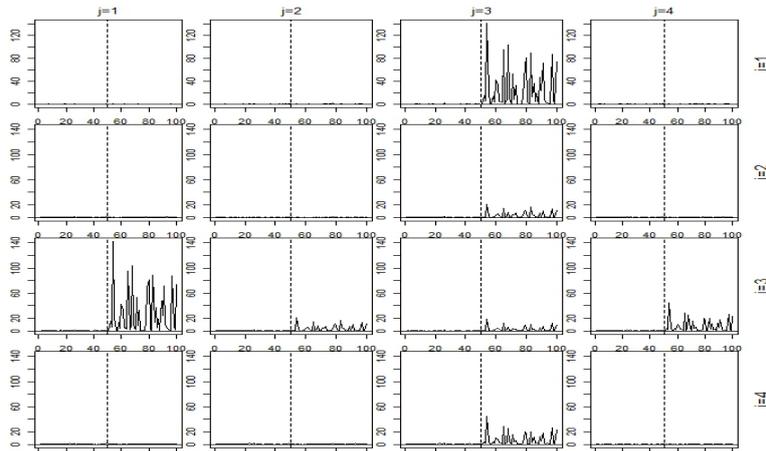


Figure 6.6. The changes in the individual-parameter contribution according to the assignable cause OC3

eter) 개별모수 기여도 R_i^I 는 i 번째 모수만을 사용한 마할라노비스 거리를 의미하고, 두 개의 모수, 즉 i 와 j 번째 모수에 의한 이모수(two parameter) 개별모수 기여도 R_{ij}^I 는 i 와 j 번째 모수만을 사용한 마할라노비스 거리를 의미한다.

개별모수 기여도가 중심위치 기여도에 보완적으로 사용될 수 있음을 알아보기 위해 이를 앞절의 예에 적용해 보자. 세 종류의 이상원인에 따른 중심위치 기여도는 Figure 6.4, Figure 6.5, Figure 6.6과 같다. 각 그림의 (i, j) 번째 그림은 R_{ij}^I 이다. Figure 6.4에서 OC1의 경우에는 개별모수 기여도가 중심위치 기여도와 마찬가지로 R_1^I 뿐만 아니라 그와 관련된 $R_{12}^I, R_{13}^I, R_{14}^I$ 가 다 같이 커지는 경향을 나타내지만, 첫 번째 모수와 직접적 관련이 없는 나머지 개별모수 기여도는 두드러지게 커짐을 보이지 않기 때문에 구체적으로 첫 번째 모수와 관련하여 변화가 발생하였음을 알 수 있다. Figure 6.5에서는 이상원인 OC2인 경우인데 유독 가운데 네 개의 그림(즉, (2, 2), (2, 3), (3, 2), (3, 3))만이 거의 변화를 보이지 않아 이 역시 첫 번째 모수와 네 번째 모수가 공동으로 마할라노비스 거리를 크게 하는데 기여하고 있음을 암시하

고 있다. Figure 6.6에서는 이상원인 OC3의 경우인데 $i = 3$ 인 경우 또는 $j = 3$ 인 경우에만 다른 그림에 비해 변화를 보이고 있어 세 번째 모수의 기여도가 유의함을 나타내고 있다. 이와 같이 개별모수 기여도도 변화를 유발한 모수와 관계가 없는 모수에 대해서는 기여도가 커지지 않아 중심위치 기여도와 함께 사용하면 원인 모수를 탐지하는데 더 효과적이라고 판단된다.

7. 다변량 T^2 관리도와의 유사성 및 차이점

마할라노비스-다구찌 시스템을 이용한 공정 모수의 재조정 절차는 다변량 T^2 관리도와 동일한 통계량을 사용함을 알 수 있다. 다변량 T^2 관리도의 목적은 평균을 검색하는 절차이지만 마할라노비스-다구찌 시스템은 특정 모수의 비교가 아닌 다변량 분포가 정상그룹에 해당되는지를 판단하는 것이다. 목적에 관계없이 제 5절의 예를 통해 마할라노비스 거리는 평균의 변화에 민감하면서도 일치적(consistent)임을 알 수 있다. 다변량 T^2 관리도는 Alt (1985) 이후 Lowry와 Montgomery (1995)와 Jensen 등 (2006) 이외에도 많은 연구가 되어오고 있다.

동일한 통계량을 사용하여도 두 절차의 사용환경은 사뭇 다르다. 먼저 관리도에서는 공정 출력치를 대상으로 하지만 재조정에서는 공정의 입력모수가 대상이다. 데이터의 관측시점을 기준하면 관리도는 피드백 관리이고 재조정은 피드포워드 관리이다. 관리도에서는 데이터의 빈번한 측정이 어려울 수 있지만 재조정은 상대적으로 빈번하여 짧은 시간에 많은 데이터를 얻을 수 있다. 관리도에서는 오경보를 줄이기 위해 제 1종 오류를 아주 작은 값이 되게 관리한계를 설정해야 하지만 재조정에서는 불필요한 재조정도 심각한 문제를 야기하지는 않기 때문에 제 1종 오류를 어느 정도 큰 값으로 설정할 수 있어 제 2종 오류를 줄일 수 있다. 어느 정도 큰 값의 제 1종 오류는 정상그룹으로부터 원하는 분위수를 설정할 수 있고 따라서 모수의 분포에 무관한(distribution-free) 경계값을 설정할 수 있다. 공정모수에 대한 다변량 분포는 알려지지 않은 경우가 많고 정규분포를 가정했을 때 위배 시 특성이 심각하게 다를 수 있다는 문제점이 있다. 이를 고려하면 분포무관 경계값을 설정할 수 있다는 점은 이 재조정 절차의 큰 장점에 해당한다고 할 수 있다.

8. 결론

마할라노비스-다구찌 시스템은 마할라노비스 거리를 사용하여 정상그룹으로 방향성 없이 동 떨어진 형태를 나타내는 비정상 그룹을 구분하는 하나의 분류방법이다. 이 논문에서는 마할라노비스-다구찌 시스템의 이런 특성을 공정 모수의 재조정 문제에 적용하는 방법을 제안하였다. 또한 공정모수의 재조정시에는 모든 모수를 동시에 조정하는 것 보다 이상원인을 유발하는데 주된 역할을 하는 하나 또는 두 개의 모수를 우선 조정하기를 제안하였고 이를 위해 기여도를 연구하였다. 기존의 기여도 연구는 모두 한 변수의 기여도만을 고려하였으나 여기서는 두 변수간의 상관관계가 존재함을 고려하여 두 변수에 의한 기여도를 고려하였다. 뿐만 아니라 차원축소에 의하면 제외하는 모수와 남아있는 모수간의 상관관계를 고려할 수 없는 문제에 착안하여 모수를 제외하는 대신 그 모수의 중심 값으로 대체하는 중심위치 기여도를 고려하였으며 결과적으로 마할라노비스 거리를 크게 하는데 주된 역할을 하는 모수의 식별에 향상된 효과가 있음을 보였다. 또한 모수의 식별에 개별모수 기여도를 같이 사용하여 원인모수의 탐지능력을 증가시킬 수 있다.

일반적인 분류문제에서는 정상그룹과 비정상그룹의 데이터가 모두 가용할 때에만 분류기준을 설정할 수 있지만 이 논문에서는 마할라노비스 거리의 분포무관 분포를 사용하여 재조정 경계값의 설정 방법을 제시하였다.

기여도에 대해서는 다양한 조건(공분산 구조, 이상원인의 다양성 등)에 대해 앞으로 더 면밀한 연구가

필요하며 특히 데이터의 분포에 따른 특성은 이 방법이 분포무관하게 사용될 수 있는지의 여부를 판단할 수 있는 중요한 연구가 되어 추후 연구로 합당하다고 판단된다. 표본 크기가 $n = 1$ 인 경우에는 Tracy 등 (1992)과 Williams 등 (2006)이 제안한 방법을 사용하여 마할라노비스-다구찌 시스템을 수정하여 사용할 수 있으며 이에 대한 연구도 추후 연구과제로 돌린다. 특히 비용문제를 고려한 경제적 재조정절차의 설계는 관리도의 경제적 설계와 유사하게 연구될 수 있으며 공정 엔지니어들에게는 통계적 설계보다 더 필요한 연구가 된다.

부록

정상그룹의 마할라노비스 거리의 순서통계량 $d_{(i)}$ 에 대해 표본분포함수는 수정항을 사용하여

$$F_m(d_{(i)}) = \frac{i}{m} - \frac{1}{2m}, \quad i = 1, 2, \dots, m$$

이라 정의한다. 상위 α 분위수를 결정하기 위해서는 표본분포함수의 값 $(1 - \alpha) + 1/(2m)$ 을 포함하는 두 순서통계량을 찾아 보간법을 사용한다. 편리상 다음을 정의한다.

$$g = \left\lceil m \left(1 - \alpha + \frac{1}{2m} \right) \right\rceil$$

표본분포함수로부터

$$F_m(d_{(g)}) = \frac{g - 0.5}{m} \leq 1 - \alpha < F_m(d_{(g+1)}) = \frac{g + 0.5}{m}$$

의 관계를 알고 보간법에 의해 마할라노비스 거리의 상위 α 분위수는

$$T_\alpha = d_{(g)} + m (d_{(g+1)} - d_{(g)}) \left(1 - \alpha - \frac{g - 0.5}{m} \right)$$

을 설정한다.

References

- Alt, F. B. (1985). Multivariate quality control, *Encyclopedia of Statistical Sciences*, **6**, N.L. Johnson and S. Kotz, (eds.) Wiley, New York.
- Jensen, W. A., Jones-Farmer, L. A., Champ, C. W. and Woodall, W. H. (2006). Effects of parameter estimation on control chart properties: A literature review, *Journal of Quality Technology*, **38**, 95–108.
- Kanetaka, T. (1988). Application of Mahalanobis distance, standardization and quality control, *Japanese Standards Association*, **41(5 and 6)**.
- Lowry, C. A. and Montgomery, D. C. (1995). A review of multivariate control charts, *IIE Transactions*, **27**, 800–810.
- Park, C. (2012). Economic design of charts when signals may be misclassified and the bounded reset chart, *IIE Transactions*, To be published.
- Park, C. (2013). An economic design of the bounded reset chart using the integral controller for batch-oriented processes, *Communications in Statistics-Theory and Methods*, DOI: 10.1080/0740817X.2012.695101.
- Runger, G. C., Alt, F. B. and Montgomery, D. C. (1996). Contributors to a multivariate statistical process control signal, *Communications in Statistics - Theory and Method*, **25**, 2203–2213.
- Taguchi, G., Chowdhury, S. and Wu, Y. (2005). *Taguchi's Quality Engineering Handbook*, John Wiley & Sons, Inc., ASI Consulting Group, LLC, Livonia, Michigan.
- Taguchi, G. and Jugulum, R. (2002). *The Mahalanobis-Taguchi Strategy*, John Wiley & Sons, Inc., New York.

- Tracy, N. D., Young, J. C. and Mason, R. L. (1992). Multivariate control charts for individual observation, *Journal of Quality Technology*, **24**, 88–95.
- Williams, J. D., Woodall, W. H., Birch, J. B. and Sullivan, J. H. (2006). Distribution of Hotelling's T^2 statistic based on the successive difference estimator, *Journal of Quality Technology*, **38**, 217–229.
- Woodall, W., Koudelik, R., Tsui, K.-L., Kim, S. B., Stoumbos, Z. and Carvounis, C. (2003). A review and analysis of the Mahalanobis-Taguchi system, *Technometrics*, **45**.