

研究論文

로버스트 추정에 근거한 수정된 다변량 T^2 -관리도

성 응 현 · 박 동 련

한신대학교 정보통계학과

Modified Multivariate T^2 -Chart based on Robust Estimation

Oong-Hyun Sung · Dong-ryeon Park

Department of Statistics and Information, Hanshin University

Abstract

We consider the problem of detecting special variations in multivariate T^2 -control chart when two or more multivariate outliers are present. Since a multivariate outlier may reflect slippage in mean, variance, or correlation, it can distort the sample mean vector and sample covariance matrix. Damaged sample mean vector and sample covariance matrix have difficulty in examining special variations clearly. An alternative to detection outliers or special variations is to use robust estimators of mean vector and covariance matrix that are less sensitive to extreme observations than are the standard estimators $\bar{\mathbf{x}}$ and \mathbf{S} . We applied popular minimum volume ellipsoid(MVE) and minimum covariance determinant(MCD) method to estimate mean vector and covariance matrix and compared its results with standard T^2 -control chart using simulated multivariate data with outliers. We found that the modified T^2 -control chart based on the above robust methods were more effective in detecting special variations clearly than the standard T^2 -control chart.

1. 서론

일변량 X -관리도는 일변량에 대한 공정 평균 μ 를 모니터하기 위해서 사용되는 관리도이다. 그러나 공정에서 얻어지는 자료가 다변량에서 얻어지는 시계열자료라면 일반적

으로 Mahalanobis 거리에 근거한 T^2 -관리도와 같은 다변량 관리도를 사용하게 된다. 관리도에서는 공정의 관리상태를 판정하기 위해서 관리한계선을 사용하며, 각 시점에 해당되는 값들이 관리한계선내에 존재하면 공정이 관리상태에 있다고 판정하게 된다.

시계열에 따라 관측된 자료가 다변량자료라면 일변량 관리도를 개별적으로 작성하기보다는 다변량 관리도를 이용하는 것이 바람직하다. 왜냐하면 p 개 변량에 대해서 개별적인 일변량 관리도를 작성할 때 가정은 p 개 변량이 서로 독립적인 관계에 있다는 것이다. 그러나 p 개 변량이 서로 상관관계가 유의하게 존재하는 경우에는 p 개 변량의 분산과 공분산을 고려한 공분산행렬(covariance matrix)을 고려한 다변량 관리도를 사용하게 된다.

일변량 관리도에서는 각 시점에 해당되는 값들이 관리한계선 밖에 존재하게 되면 그 시점에서는 공정이 이상원인에 의한 것으로 판정하게 되고, 다변량 T^2 -관리도에서는 개별 관측벡터와 표본평균벡터사이의 통계적 거리(statistical distance)의 제곱이 관리한계선 안에 존재하면 공정이 관리상태에 있다고 판정하게 된다. 이러한 절차는 자료에서 통계적으로 이상값(outlier)을 탐지하는 절차와 같다. 일변량자료에서는 두개 혹은 여러 개의 이상점이 존재할 경우에는 평균(mean)과 분산(variance)의 팽창으로 인해서 상대적으로 약한 극단값(extreme values)들을 탐지하는데 어려워지는 경우가 발생하게 된다. 반면에 p 개 변량이 서로 선형적인 관계가 유의하게 존재하는 다변량자료에서 다변량 이상값이 존재하는 경우에는 평균과 분산뿐만 아니라 공분산(covariance) 혹은 상관(correlation)에 심각한 영향을 미치는 경우가 발생하게 된다.

다변량 T^2 -관리도에서 사용되는 통계적 거리는 p 개 변량의 공분산행렬을 사용하여 구하게 된다. 이때 공분산행렬은 다변량 이상값에 의해서 왜곡될 수 있기 때문에, 공분

산행렬을 고려한 통계적 거리는 다변량자료에서 이상값인 특별변동을 탐지하는데 문제가 발생된다. 다변량자료에서 이상값이 공분산행렬에 상당한 영향을 미치는 경우에는 공분산행렬의 추정에 영향을 최소화시킬 수 있는 로버스트 방법(robust method)을 사용하는 것이 바람직 할 것이다. 다변량 이상값들이 상관구조에 영향을 미치는 경우에는 MVE(minimum volume ellipsoid)방법 혹은 MCD(minimum covariance determinant)방법에 추정된 공분산행렬을 추정하는 것이 바람직하다고 판단된다. 공분산행렬추정에 사용되는 두가지 로버스트방법은 아직까지 다변량 품질관리도에서 특별변동을 탐지하는데 적용된 예가 없기 때문에, 본고에서는 로버스트방법을 이용하여 일반적으로 사용되는 다변량 T^2 -관리도의 문제점을 개선하고자 한다.

2. 다변량 T^2 -관리도와 이상값 탐지

2.1 다변량 T^2 -관리도와 일변량 X -관리도

평균벡터가 μ 이고 공분산행렬이 Σ 인 p -변량 다변량분포에서 얻어진 관측벡터를 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ 라고 했을 때 μ 와 Σ 의 추정량인 표본 평균벡터와 표본 공분산행렬을 $\bar{\mathbf{x}}$ 와 \mathbf{S} 라고 하자. 이때 j -번째 관측벡터 \mathbf{x}_j 와 표본 평균벡터 $\bar{\mathbf{x}}$ 사이의 통계적거리의 제곱을 T_j^2 로 표시하면

<표 1> 이변량 생성자료

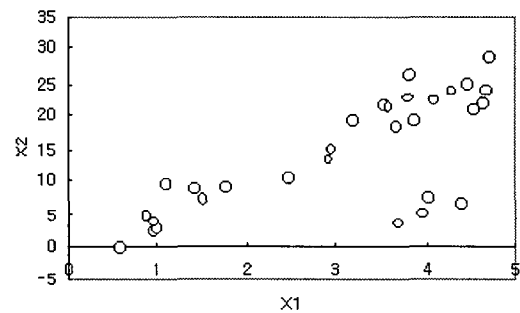
번호	1	2	3	4	5	6	7	8	9	10
X_1	3.81	1.42	1.12	0.96	4.63	3.69	4.02	1.01	1.51	2.48
X_2	22.60	8.83	9.50	2.04	21.37	18.24	7.47	2.53	7.20	10.36
번호	11	12	13	14	15	16	17	18	19	20
X_1	0.97	2.91	0.59	3.52	3.95	3.58	0.87	4.66	4.30	4.70
X_2	3.65	13.02	-0.34	21.33	5.03	20.93	4.74	23.97	23.90	28.56
번호	21	22	23	24	25	26	27	28	29	30
X_1	3.18	1.78	2.96	3.87	3.70	3.82	4.08	4.48	4.54	4.42
X_2	19.10	8.92	14.94	19.09	3.38	26.21	22.19	24.95	20.61	6.60

$$T_j^2 = (\mathbf{x}_j - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}) \quad (1)$$

이 된다. 모든 T_j^2 값들을 축에 따라 하한계를 $LCL=0$, 상한계를 $UCL = \chi_{\alpha, p}^2$ 로 하여 산점도를 작성하면 다변량 T^2 -관리도가 된다. 만약 $T_j^2 > \chi_{\alpha, p}^2$ 이면 j -번째 관측벡터 \mathbf{x}_j 는 특별변동으로 간주한다. 식(1)의 문제점은 다변량 이상값의 위치에 따라 표본공분산행렬의 구조를 왜곡시킬 수 있기 때문에 다변량 T^2 -관리도에서 특별변동을 탐지하는데 어려움이 발생되게 된다. 왜냐하면 다변량 이상값은 변수들의 평균, 분산 혹은 상관관계에 영향을 미치기 때문이다. 본고에서는 다변량 이상값을 시각적으로 표현하기 위해서 변수의 수가 $p=2$ 인 경우에 대해서 살펴보기로 하자.

변수의 수가 $p=2$ 인 경우에 자료를 생성하기 위해서 첫 번째 변수 X_1 의 자료는 범위가 $0 < X_1 < 5$ 인 일양분포 (uniform distribution)에서 난수를 $n=26$ 개 발생시키고, 두 번째 변수 X_2 자료는 $X_2 = 1 + 5X_1 + \varepsilon$,

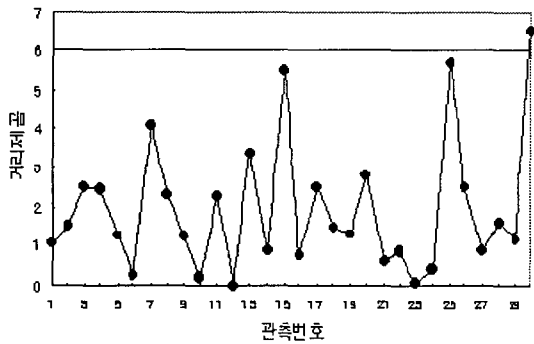
$\varepsilon \sim N(0, 2)$ 를 통해서 자료를 생성하였다. 또한 상관관계에 영향을 미칠 수 있는 이상값 4개를 추가하였는데, X_1 은 범위가 $3.5 < X_1 < 4.5$ 인 일양분포에서, X_2 는 범위가 $0 < X_2 < 10$ 인 일양분포에서 각각 생성하였다. 위 절차에 의해서 생성된 자료는 아래 <표 1>과 같다. 또한 생성된 이변량 자료에 대해서 산점도를 작성해보면 <그림 1>과 같다.



<그림 1> 이변수 산점도

<그림 1>에 의하면 우측 하단에 있는 4개의 관측자료(관측번호 7, 15, 25, 30 번째)는 두 변수사이의 상관에 영향을 미칠 수 있

는 이상값이 될 가능성이 있다. 두 변수사이의 표본상관계수는 $n=26$ 인 경우에는 $r_{12} = 0.96$ 이고, 이상값들을 포함한 $n=30$ 인 경우에는 $r_{12} = 0.74$ 이다. 상관구조에 영향을 미치는 이상값들이 존재할 경우 다변량 T^2 -관리도와 일변량 X -관리도에서 특별변동으로 분명하게 탐지할 수 있는지를 검토하여보자.



<그림 2> 다변량 T^2 -관리도

< 표 1 >의 이변량자료에 근거해서 식(1)을 이용하여 작성된 T^2 -관리도는 < 그림 2 >와 같다. T^2 -관리도는 유의수준 $\alpha = 0.05$ 에서 하한계는 $LCL=0$ 이고 상한계는 $UCL = \chi^2_{0.05, 2} = 5.99$ 이다.

다변량 T^2 -관리도에 의하면 $j=30$ 인 T^2_{30}

을 제외한 모든 기간에 걸쳐서 T^2_j 는 관리한계내에 위치하기 때문에, 다변량 자료에서 유의한 특별변동은 $j=30$ 에서 발생되었다. 그리고 상한계에 상대적으로 가까운 T^2_j 순으로는 $T^2_{25}, T^2_{15}, T^2_7$ 으로 나타났다.

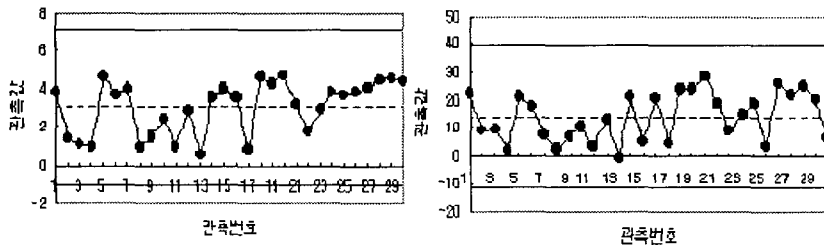
< 그림 2 >에 의하면 이변량 자료에서 구한 다변량 T^2 -관리도는 종합적으로 특별한 변동이 거의 없는 안정되어있는 것으로 판정하게 된다.

< 그림 2 >의 다변량 T^2 -관리도에서는 T^2_{30} 를 제외한 모든 기간에서 관리한계안에 위치하고 있다.

또한 일변량 자료에서 특별 변동이 존재하는지를 검토하기 위해서는 일변량 관리도인 X -관리도를 개별적으로 작성하여야 한다. 이때 개별 변수에 대한 X -관리도에서 하한계와 상한계를 아래와 같이 설정하게 된다.

$$LCL = \bar{X}_i - 3s_i, UCL = \bar{X}_i + 3s_i \quad (2)$$

식(2)를 이용하여 작성된 < 그림 3 >의 일변량 두개 X -관리도에 의하면 모든 관측값들이 관리상한계 안에 위치하고 있기 때문에 특별변동은 없는 것으로 판정된다. 다변량 T^2 -관리도에서 T^2_{30} 가 특별변동으로 판정되었으나 일변량 X -관리도에서는 특별변동이 관측되지 않은 이유는 두 변수사이의 상



<그림 3> 일변량 X -관리도

관관계를 무시하여 작성되었기 때문이다.

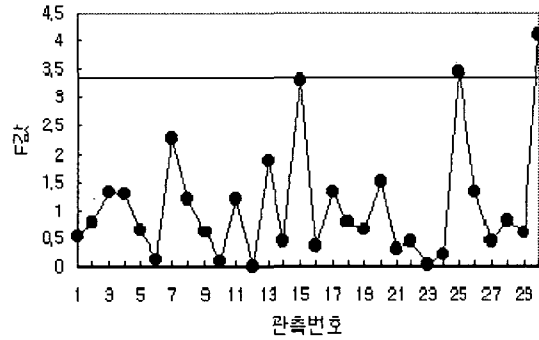
2.2 다변량 이상값 F-검정

다변량 이상값을 탐지하기 위한 대표적인 방법으로는 Wilks(1963), Mardia(1970), Bacon-Shone 과 Fung(1987), Yang 과 Lee(1987) 등이 있는데, 본절에서는 다변량 이상값을 효율적으로 탐지하기 위해서 Yang 과 Lee(1987)이 제안한 통계량인

$$F_j = \frac{n-p-1}{p} \left[\frac{1}{1-nT_j^2/(n-1)^2} - 1 \right],$$

$$j = 1, 2, \dots, n. \tag{3}$$

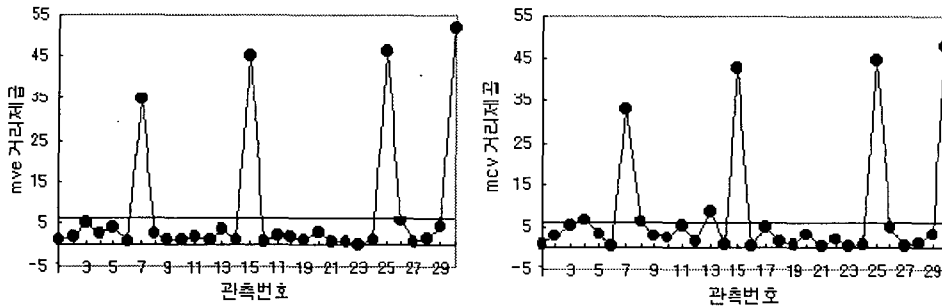
을 이용해서 분석하기로 하자. 식(3)은 분자의 자유도가 p 이고 분모의 자유도가 $n-p-1$ 인 $F_{p, n-p-1}$ 분포에 따르게 된다. 식(3)에서 구해진 F_j 값들은 유의수준 0.05에서 $F_j \geq F_{0.05, p, n-p-1}$ 인 경우에 대응되는 j -번째 관측벡터 \mathbf{x}_j 는 다변량 이상값으로 판정하게 된다. < 표 1 >의 자료에 대하여 식(3)을 이용하여 구한 F_j 값들과 기각치 $F_{0.05, 2, 27} = 3.354$ 와 비교해서 귀무가설 (H_0 : j -번째 관측벡터 \mathbf{x}_j 는 다변량 이상값이 아니다)의 기각여부를 결정하게 된다. 기각치를 상한선을 하여 다변량 차트를 작성하면 < 그림 4 >와 같다. < 그림 4 >에 의하면 F_{25} 와 F_{30} 은 다변량 이상값으로 판정되었고 $F_{15} = 3.293$ 으로 상한선 경계에 가깝게 나타났다. < 그림 4 >와 < 그림 2 >를 비교하면 일반적인 다변량 T^2 -관리도에서 탐지하지 못한 25번째 관측벡터를 특별변동으로 탐지할 수 있었으나, 관리상한선에서 약간 벗어나 특별변동들을 분명하게 탐지하고 있는 못하고 있다.



<그림 4> 다변량 이상값 탐지

3. 로버스트거리에 대한 수정된 T^2 -관리도

일변량 X -관리도를 작성하는데 사용된 식(2)을 살펴보면 표본평균 \bar{X}_i 와 표본표준편차 s_i 는 이상값에 의해서 영향을 많이 받을 수 있기 때문에 관리 하한계와 상한계의 차이가 넓어지게 되어 추가적인 특별변동을 탐지하는데 문제가 발생할 수 있다. 따라서 일변량자료에서는 두개 혹은 여러 개의 이상점이 존재할 경우에는 평균(mean)과 분산(variance)의 팽창으로 인해서 상대적으로 약한 극단값(extreme values)들을 탐지하는데 어려워지는 경우가 발생된다. 더구나 p 개 변량에서 관측된 다변량자료에서 다변량 이상값이 존재하는 경우에는 평균과 분산뿐만 아니라 상관(correlation)에 영향을 미치는 경우가 발생하게 되어 평균벡터와 표본공분산행렬을 왜곡시키게 된다. 따라서 다변량 이상값들이 존재하는 경우 식(1)을 이용하여 다변량 T^2 -관리도를 작성할 때에는 평균벡터와 공분산행렬을 다변량 이상값에 민



<그림 5> 이변량 로버스트 T^2 -관리도

감하지 않는 로버스트방법에 의해서 추정하는 것이 바람직할 것이다. 로버스트 공분산 행렬을 추정하기 위한 방법에 대해서는 Campbell(1980), Huber(1981), Lopuhaa(1989), Maronna(1976), Rousseeuw와 Leroy(1987) 등이 논의하였고, 본 절에서는 가장 널리 사용되고 방법으로는 Rousseeuw와 Leroy(1987)가 제안한 MVE 방법과 MCD 방법을 다변량 T^2 -관리도를 작성하는데 사용하도록 하겠다.

MVE 방법에 의하여 추정되는 표본평균벡터 $\bar{\mathbf{x}}^*$ 는 다음과 같이 정의된다.

$\bar{\mathbf{x}}^*$ = n 개 p -변량 관측자료 중에서 적어도 h 개 관측자료를 포함하는 체적(volume)이 최소인 타원체(ellipsoid)의 중심 일반적으로 h 는 자료 개수의 절반인 $[n/2]+1$ 이 된다. 단, $[n/2]$ 는 $n/2$ 의 정수 부분을 나타낸다. 표본공분산행렬은 $\bar{\mathbf{x}}^*$ 를 추정하는 과정에서 얻어진 타원체에 적절한 상수를 곱하여

\mathbf{S}^* 로 추정된다. MVE 방법을 실제 적용시키는 과정에서 자료의 절반에 대한 모든 경우를 고려하여 그것을 포함하는 최소 타원체의 체적을 계산하는 것은 극히 비효율적인 작업이 된다. 따라서 일반적인 경우에는 근

사적인 알고리즘에 의존하게 되며, 이 방법에 대한 자세한 내용은 Rousseeuw and Leroy(1987)에 있다. MVE 방법에 의해서 추정된 표본평균벡터와 표본공분산행렬을 식(1)에 대입하여 구한 식은 아래와 같다.

$$T_j^{2*} = (\mathbf{x}_j - \bar{\mathbf{x}}^*)' \mathbf{S}^{*-1} (\mathbf{x}_j - \bar{\mathbf{x}}^*) \quad (4)$$

MVE 추정량과는 다른 접근 방법으로 로버스트 추정량을 계산하는 방법으로 MCD 추정 방법이 있다.

이 방법은 로버스트 회귀분석에서 사용되는 Least Trimmed Squares(LTS) 추정량의 개념을 일반화시킨 것으로 MCD에 의하여 추정되는 표본평균벡터는 $\bar{\mathbf{x}}^{**}$ 는 다음과 같이 정의된다.

$\bar{\mathbf{x}}^{**}$ = n 개 p -변량 관측자료 중에 공분산행렬의 행렬식(determinant)이 최소가 되는 h 개 관측자료들의 표본평균벡터

이 방법은 결국 표본공분산행렬에 의하여 타원체를 구할 경우, 체적이 최소가 되도록 하는 h 개 자료를 찾아서 그 자료의 중심을 계산하는 것이 된다. MCD 방법에 의해서 추정된 표본평균벡터와 표본공분산행렬 \mathbf{S}^{**} 을 식(1)에 대입하여 구한 식은 아래와 같다.

$$T_j^{2**} = (\mathbf{x}_j - \bar{\mathbf{x}}^{**})' \mathbf{S}^{**^{-1}} (\mathbf{x}_j - \bar{\mathbf{x}}^{**}) \quad (5)$$

MVE 와 MCD 방법은 전체 자료 중에서 절반까지 이상값에 의하여 오염이 되더라도 평균벡터와 공분산행렬을 로버스트하게 추정할 수 있는 방법으로 high breakdown point 를 갖는 추정량인 것으로 알려져 있다. 두 가지 로버스트 추정량에 의해서 수정된 이변량 로버스트 T^2 -관리도는 < 그림 5 >와 같다. MVE 와 MCD 방법을 사용한 로버스트 T^2 -관리도는 거의 같은 변동 형태를 보이고 있다. < 그림 5 > 의 두 가지 로버스트 T^2 -관리도에서는 <그림 2 >의 일반적인 이변량 T^2 -관리도에서 특별변동으로 탐지하지 못한 7, 15, 25 번째의 관측벡터를 뚜렷하게 보여주고 있다. 따라서 서로 상관관계가 있는 다변량자료에서 관측된 특별변동(이상값)들이 여러 개 존재하는 경우에는 표본평균벡터와 표본공분산행렬을 심각하게 왜곡시킬 수 있기 때문에 일반적인 T^2 -관리도보다는 로버스트 평균벡터와 공분산행렬 추정에 근거한 수정된 T^2 -관리도를 사용하는 것이 바람직하다고 판단된다.

4. 적용예제

예제들을 통하여 MCD와 MVE 에 의하여 수정된 T^2 -관리도에 의하여 이상값을 탐지하는 방법의 효율성을 비교하여 보자. 자료를 생성하는 모형으로 다음의 모형들이 사용되었다.

1) 모형 1 :

$$X_{1i} = 1 + 5X_{2i} + \epsilon_i, \quad i=1, \dots, 50$$

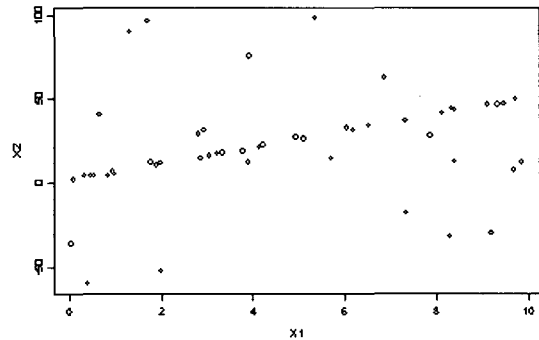
$$X_{2i} \sim \text{Unif}(0, 10),$$

$$\epsilon_1, \dots, \epsilon_{30} \sim N(0, 1), \quad \epsilon_{31}, \dots, \epsilon_{50} \sim N(0, 50)$$

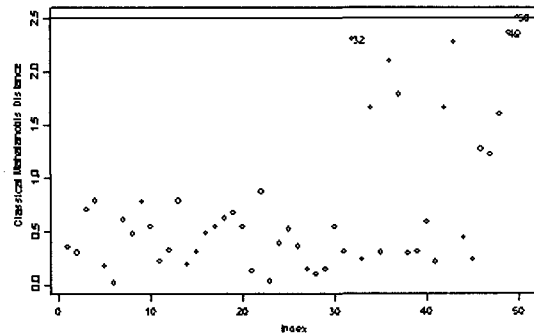
2) 모형 2 :

$$X_{1i} = 1 + 5X_{2i} + \epsilon_i, \quad i=1, \dots, 50$$

$$X_{2i} \sim \text{Unif}(0, 1), \quad \epsilon_i \sim \text{Cauchy}(0, 0.5)$$

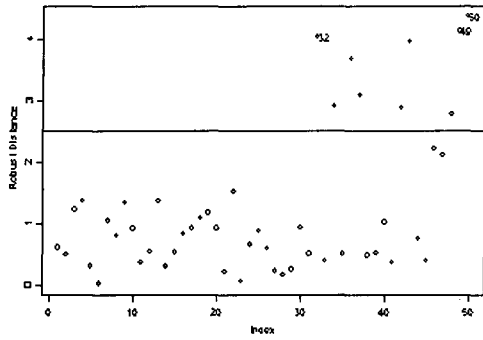


<그림 6> 모형 1의 산점도

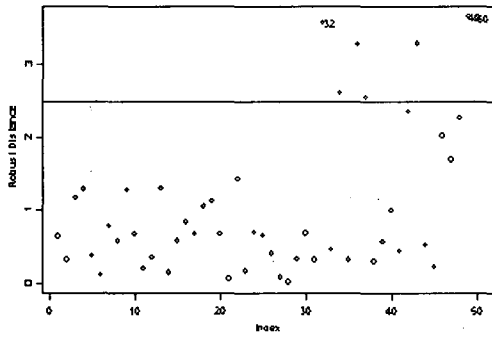


<그림 7> T^2 -관리도

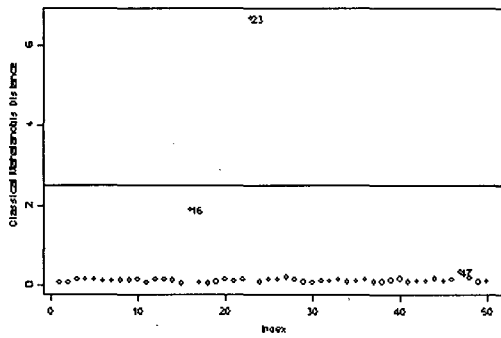
모형 1에서는 분산이 50인 정규분포에서 생성된 자료 중에 이상값으로 판정되어야 할 자료가 다수 포함되어있을 것으로 예측된다. < 그림 6 >에는 (X_1, X_2) 의 산점도가 그려져 있고 < 그림 7 >에는 일반적인 T^2 -관리도에 의한 판정결과가, < 그림 8 >에는 MVE에 의하여 수정된 T^2 -관리도에 의한



<그림 8> MVE에 의한 로버스트 T^2 -관리도



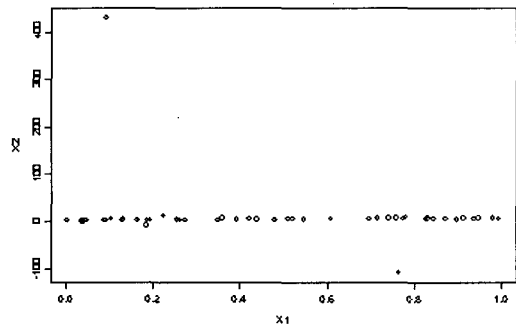
<그림 9> MCD에 의한 로버스트 T^2 -관리도



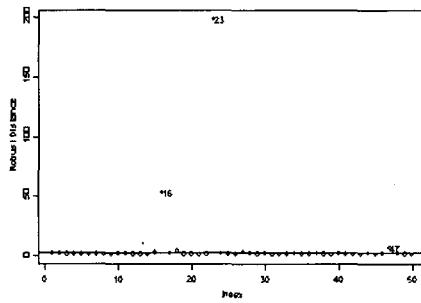
<그림 10> T^2 -관리도

판정 결과가, < 그림 9 >에는 MCD에 의하여 판정된 결과가 각각 그려져 있다. < 그림 7 >의 일반적인 T^2 -관리도에서는 이상

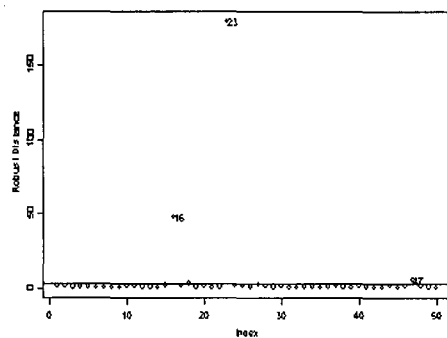
값을 전혀 인식하지 못하고 있으나 < 그림 8 >과 < 그림 9 >의 로버스트 T^2 -관리도에서는 다수의 이상값을 정확하게 식별하고 있음을 확인할 수 있다.



<그림 11> 모형 2의 산점도



<그림 12> MVE에 의한 로버스트 T^2 -관리도



<그림 13> MCD에 의한 로버스트 T^2 -관리도

모형 2의 자료에 대한 산점도는 <그림 10>에 있다. 산점도에서 확인 할 수 있듯이 두 개의 명백한 이상값이 존재하는 자료이다. 그러나 <그림 11>의 일반적인 T^2 -관리도에는 Masking 효과에 의하여 한 개의 이상값만이 인식되고 있다. <그림 12>와 <그림 13>의 로버스트 T^2 -관리도에는 두 개의 이상값이 인식되고 있는 것을 알 수 있다.

5. 결론

재화나 서비스의 품질을 향상시키기 위해서는 공정과정에서 수집된 자료로부터 특별변동을 탐지하고 그 원인을 조사하여야 한다. 공정과정에서 수집되는 자료는 일변량인 경우보다는 다변량인 경우가 대부분이다. 공정과정에서 수집되는 다변량자료는 서로 독립적이라기 보다는 서로 상관이 있는 자료가 대부분이다. 다변량자료에서 일변량 개별 관리도를 작성할 경우에는 변수들 사이의 상관구조를 무시되고, 일변량 관리도에서 탐지되지 않은 특별변동도 다변량 관리도에서 탐지될 수 있는 경우가 발생되게 된다. 반대로 일변량 관리도에서 탐지된 특별변동은 다변량 관리도에서는 특별변동으로 간주할 수 없는 경우가 흔히 발생되게 된다.

다변량자료에서 발생된 두 개 혹은 여러 개 특별변동이 존재할 경우에는 변수들의 평균과 분산 뿐만 아니라 상관(correlation)에 영향을 미치는 경우가 발생하게 되어 표본평균벡터와 표본공분산행렬을 왜곡시키게 된다. 따라서 다변량 T^2 -관리도를 작성할 때에는 표본평균벡터와 공분산행렬을 다변량 이상값에 민감하지 않는 로버스트방법에 의

해서 추정하는 것이 바람직할 것이다. 본고에서는 Rousseeuw 와 Leroy(1987)가 제안한 MVE 방법과 MCD 방법을 사용하여 수정된 로버스트 다변량 T^2 -관리도를 작성한 결과 다변량 이상값인데도 불구하고 일반적인 다변량 T^2 -관리도에서 탐지하지 못한 특별변동을 뚜렷하게 탐지할 수 있었다. 따라서 서로 상관관계가 유의하게 존재하는 다변량 자료에서 두 개 혹은 여러 개 이상값이 존재할 경우에는 본고에서 제안한 수정된 로버스트 다변량 T^2 -관리도를 작성하는 것이 바람직 할 것으로 판단된다.

참고문헌

- [1] Bacon-Shone, J., and W. K. Fung (1987), "A New Graphical Method for Detecting Single and Multiple Outliers in Univariate and Multivariate Data," *Applied Statistics*, Vol. 36, no. 2, pp 153-162.
- [2] Campbell, N. A. (1980), "Robust Procedures in Multivariate Analysis I: Robust Covariance Estimation," *Applied Statistics*, Vol. 29, pp. 231-237.
- [3] Huber, P. J. (1981), *Robust Statistics*, New York: Wiley.
- [4] Lopuhaa, H. P. (1989), " On the Relation between S-estimators and M-estimators of Multivariate Location and Covariance," *Annals of Statistics*, Vol. 17, pp. 1662-1683.
- [5] Mardia, K. V. (1970), " Measures of Multivariate Skewness and Kurtosis

- with Applications," *Biometrika*," Vol. 57, pp.519-530.
- [6] Maronna, R. A. (1976), " Robust M-estimators of Multivariate Location and Scatter," *Annals of Statistics*, Vol. 4, pp.51-67.
- [7] Rousseeuw, P. J, and Leroy, A. M. (1987), *Robust Regression and Outlier Detection*, New York: Wiley.
- [8] Wilks, S. S. (1963), "Multivariate Statistical Outliers," *Sankhya, Series A*, Vol. 25, pp. 407-426.
- [9] Yang, S. S., and Lee, Y. (1987), "Identification of a Multivariate Outlier," Presented at the Annual Meeting of the American Statistical Association, San Francisco, CA, August 1987.