

상관계수의 변동을 탐지하기 위한 EWMA 관리도

한정혜

충북대학교 대학원 전자계산학과

조중재

충북대학교 통계학과

EWMA Control Chart for Monitoring a Process Correlation Coefficient.

Jeong-Hye Han

Dept. of Computer Science, Chungbuk National University

Joong-Jae Cho

Dept. of Statistics, Chungbuk National University

Abstract

The EWMA(Exponentially Weighted Moving Average) has recently received a great deal of attention in the quality control literature as a process monitoring tool on the shop floor of manufacturing industries, since it is easy to plot, to interpret, and its control limits are easy to obtain. Most applications of the EWMA for process monitoring have concentrated on the problem of detecting shifts of a process mean and a process standard deviation with ARL(Average Run Length) properties. But there may be the necessity of controlling linearity on product quality such as the correlation coefficient to the process operator. Control managers may want to protect the increase of a process correlation coefficient value, such as 0, between two variables of interest. However, there are few studies concerned on this part. Therefore, we propose EWMA models for a process correlation coefficient using two transformed statistics, T -statistic and (Fisher's) Z -statistic. We also present some results of simulation by SAS/IML and compare two models.

1. 서론

품질관리의 중요한 목적 중 하나는 공정에 이상요인이 발생하였을 경우 이를 가능한 빨리 탐지하여 수정조치를 취함으로써 불량제품의 사전 발생을 억제하는 것이다. 이러한 목적으로 산업현장에서는 주로 관리도를 사용하는데, 요즘은 CUSUM 관리도와 EWMA 관리도가 널리 이용되고 있다. 특히 EWMA 관리도는 'Roberts(1959)'가 제안한 관리도로서, 여러 분야에 응용되고 있으며, 'Robinson & Ho(1978), Crowder(1987)', 'Crowder(1989)', 'Lucas & Saccucci(1990)'등에서는 EWMA 관리도를 ARL 관점에서 연구하였으며, 'Hunter(1986)'에는 EWMA 관리도의 ARL(Average Run Length)을 자세히 소개 되어있다. 최근에는 공정평균의 관리 뿐만 아니라 공정 표준편차의 관리에도 EWMA 관리기법이 활발히 적용되고 있다. 'Crowder & Hamilton(1992)'는 표본분산의 로그변환 $\ln S^2$ 을 이용한 EWMA모형 소개하고, Shewhart 관리도, CUSUM 관리도, EWMA 관리도의 ARL을 비교했다. 'Montgomery & Mastrangelo(1991)'은 한 공정 특성변수의 관측치들이 시간에 따라 자기상관을 가졌을 경우에 EWMA 모형을 적용한 결과와 시계열의 AR모형을 적용한 후 잔차 관리도의 결과를 비교하였다.

이와 같이 제품생산공정의 조정·감시를 위한 이러한 EWMA 관리도 기법들은 평균런길이를 이용한 공정평균의 이동을 탐지하는 문제에 집중되어왔다. 하지만, 제품의 품질에 중요한 영향을 끼칠 수 있는 공정변동의 증가뿐만 아니라 서로 상관관계가 의미 있는 두 개 이상의 변수들을 사용할 경우 두 변수의 선형성에 대한 이상증감을 보다 쉽게 실제적으로 탐지하는 것 역시 중요한 문제가 될 수 있으나 아직 연구가 미흡한 실정이다. 만일 공정관리자가 동시에 두 개 이상의 공정 특성변수 사이의 공분산이나 상관관계를 관리하고자 할 경우에 주변 특성변수만의 관리로는 부족할 것이다. 즉, 두 특성변수를 동시에 관리하고자 할 경우 각 특성변수의 주변분포에 대한 평균이나 분산을 쉽게 관리할 수 있으나, 이와 같은 관리만으로 두 특성변수의 관련성을 나타내는 측도인 공분산의 변동을 관리하는 것은 올바른 경보를 주기 위한 정보로서 충분치 않다. 이때 이용할 수 있는 측도가 공분산과 관련된 상관계수로서 이에 대한 EWMA 관리도 연구인 '장덕준(1994)'에서는 네 개의 모수 $\mu_x, \mu_y, \sigma_x, \sigma_y$ 를 아는 이변량 정규분포인 경우의 모상관계수 추정량을 이용하여 상관계수의 변동을 감지하기 위한 다변량 EWMA 관리도를 제시하였으나 이에 관한 계속적인 연구가 부족하다. 그리고 공정관리자가 두 공정 특성변수 사이의 상관계수를 관리하고자 할 대부분의 경우 위 네 개의 모수들을 알지 못하므로 실제 관측치들에서 얻을 수 있는 표본상관계수 값으로 관리해야 할 것이다.

따라서 본 논문의 2장에서는 해석의 용이성을 갖으며 다른 관리도들과 절충형인 EWMA 관리모형을 살펴보았다. 3장에서는 두 특성변수의 선형측도인 상관계수의 관리를 위하여 표본상관계수값을 변환한 T -통계량과 Z -통계량을 적용한 두 개의 EWMA 모형을 제안하였으며, 이 두 모형에 대한 관리한계도 제시하였다. 그리고 4장

에서는 3장에서 제시된 EWMA 모형과 관리한계에 대하여 여러 가지 소표본이 주어졌을 때의 상관계수 변동에 이상경보가 나타나는 평균기대시간이라 할 수 있는 ARL 값과 표준편차 SDRL(Standard Deviation of Run Length)값을 시뮬레이션 하였다. 마지막 5장에서는 4장의 시뮬레이션 결과를 토대로 보다 바람직한 관리도 모형을 선택하기 위한 여러 가지 성질들을 분석·제시하였다.

2. EWMA 관리도

본 장에서는 EWMA 관리모형을 살펴보겠다. 평균이 μ_x , 분산이 σ_x^2 인 서로 독립이러 가정되는 통계량 x_t 에 의해 생성되는 EWMA_t 값은 관리될 공정모수의 사전 예측치(priori estimate)인 EWMA₀값에서 시작하며 다음과 같은 모형 식으로 표현될 수 있다.

$$EWMA_t = (1 - \lambda) EWMA_{t-1} + \lambda x_t$$

단, EWMA_t : t 시점에서의 EWMA 값

x_t : t 시점에서의 관찰값

λ : 평활모수 ($0 < \lambda \leq 1$)

이때, λ 는 EWMA의 자료의 영향의 정도를 결정하는 상수로서 평활모수(smoothing parameter)로 λ 값이 작을 수록 과거의 자료에 상대적으로 큰 가중치를 준다. 따라서 평활모수 λ 의 추정은 공정관리자에게는 중요한 문제로 반복적인 최소제곱 추정법을 사용하여 구할 수 있다. 일반적으로 λ 값의 추정에 필요한 관찰값은 적어도 50개 이상이어야 적당하다고 알려져 있으며 대개 0.05~0.25의 값을 갖는다[Hunter(1986)]. $\lambda=1$ 일 때 EWMA는 가장 최근 관측치에 가중치를 1로 주는 Shewhart 관리도와 같다. 그리고 λ 가 0에 가까울 때 EWMA는 가장 최근 관측치에 더 적은 가중치를 주어 CUSUM 관리도와 그 특성이 유사해진다. 이러한 EWMA 모형은 평균이 $\mu_{EWMA} = \mu_x$, 분산이 $\sigma_{EWMA}^2 = \sigma_x^2 \left(\frac{\lambda}{2-\lambda} \right) [1 - (1-\lambda)^{2t}] \cong \sigma_x^2 \left(\frac{\lambda}{2-\lambda} \right)$, $t \rightarrow \infty$ 인 확률분포를 갖는다. 이에 대한 자세한 내용은 'Hunter(1986)'을 참고하면 될 것이다.

3. 상관계수에 대한 EWMA 모형

본 장에서는 두 변수 X, Y 간의 선형성을 관리할 수 있는 기존의 EWMA모형과 제안된 EWMA모형들을 살펴보면 다음과 같다. $BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ 의 동일한 분포를 갖는 확률변수 $(X_t, Y_t), t=1, 2, \dots$ 가 시간이 경과함에 따라 각 시점에서 서로 독립적인 부분군의 표본크기가 n 인 관측 쌍들 $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$ 을 취한다고 가정하자. 만일 모집단의 모수를 아는 경우 '장덕준(1994)'은 상관계수를 관리하기 위한 다음의 EWMA 모형을 제시하였다.

$$\begin{aligned} EWMA_t &= (1-\lambda) EWMA_{t-1} \\ &+ \lambda \sum_{j=1}^n \frac{(x_{tj} - \mu_x)(y_{tj} - \mu_y)}{n\sigma_x\sigma_y}, \quad 0 < \lambda \leq 1, \quad t=1, 2, \dots \end{aligned} \quad (1)$$

'장덕준(1994)'에서는 식(1)에 의한 상관계수 변동을 감지하기 위한 EWMA 관리도의 수행도가 좋음을 보였으며, 특히 가변추출간격(Variable Sampling Interval)을 적용한 경우가 고정추출간격(Fixed Sampling Interval)의 경우보다 더 좋음도 보였다. 그러나 식(1)의 모형에 의해 상관계수의 변동을 관리하고자 하는 경우, 실제적으로 모수를 모르는 경우가 대부분이므로 응용하기 어렵다.

따라서 모상관계수 ρ 의 추정량인 표본상관계수 R 을 이용한 EWMA 모형을 다음 절에서 제시하고자 한다. 이때, 표본상관계수 R 은 표본 크기가 n 일 때 다음과 같이 정의된다.

$$R = \frac{\sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_{j=1}^n (x_j - \bar{x})^2} \sqrt{\sum_{j=1}^n (y_j - \bar{y})^2}}$$

3.1. 상관계수 변동의 관리를 위한 EWMA 모형

표본상관계수 R 의 분포는 'Fisher(1915)'가 처음으로 연구하였는데, 확률표본 $(X_t, Y_t), t=1, 2, \dots$ 의 분포가 이변량 정규분포인 경우조차 그 분포가 상당히 복잡한 형태를 갖는다. 그러나 모상관계수 $\rho=0$ 라고 가정하면 표본상관계수 R 의 함수인 다음의 T -통계량의 분포는 다음과 같다.

$$T = \frac{\sqrt{n-2}R}{\sqrt{1-R^2}} \sim t(n-2)$$

즉, 변환된 T 통계량의 분포는 t 분포를 따르며 n 이 클수록 표준정규분포에 근사한다. 따라서 목표치 $\rho=0$ 인 경우에 상관계수의 변화에 대하여 T -변환된 통계량을 이용한 EWMA 모형은 다음과 같이 주어진다.

$$EWMA_t = (1-\lambda) EWMA_{t-1} + \lambda T_t, \quad t=1, 2, \dots \quad (2)$$

단, $EWMA_t$: t 시점의 EWMA 값

λ : 평활모수 ($0 < \lambda \leq 1$)

$$R_t = \frac{\sum_{j=1}^n (x_{tj} - \bar{x}_t)(y_{tj} - \bar{y}_t)}{\sqrt{\sum_{j=1}^n (x_{tj} - \bar{x}_t)^2} \sqrt{\sum_{j=1}^n (y_{tj} - \bar{y}_t)^2}}$$

T_t : t 시점에서 계산된 두 변수 X, Y 의 표본상관 계수 R_t 에 대한

$$\text{변환값} \frac{\sqrt{n-2}R_t}{\sqrt{1-R_t^2}}$$

따라서 만약 두 특성치 변수들이 서로 일차 선형성이 없도록 하는 경우가 목표인 실제 공정관리 현장에서 이 상관계수값을 관리하고자 한다면, 목표치는 $\rho=0$ 인 식(2)의 EWMA 관리모형을 사용하여 R_t 값들이 0으로부터 이탈하는 것을 관리하게 될 것이다.

그러나 보다 일반적인 목표치가 $\rho \neq 0$ 인 경우에는 표본상관계수 R_t 의 확률분포를 이용하여 상관계수를 관리하고자 하는 경우가 있을 수 있으며 이 경우 T -변환된 R_t 를 사용한 식(2)의 모형을 적용할 수 없다. 그런데 관리 목표치가 $\rho = \rho_0 \neq 0$ 인 경우는 'Fisher(1921)'가 제안한 표본상관계수 R_t 의 Z -변환을 고려할 수 있을 것이다.

Z -변환한 통계량 $Z = \frac{1}{2} \log \left\{ \frac{1+R}{1-R} \right\}$ 는 표본상관계수 R 에 대하여 일대일 대응함수로 단조 증가한다. 또한 이 변환은 ρ 에 대한 신뢰구간을 제공하므로 이를 이용하여 'Hall & Martin(1989)'는 상관계수에 대한 비모수적 불스트랩 신뢰구간을 구하는데 적용하기도 했다. 이와 같은 표본상관계수 R 을 Z -변환을 시킨 통계량은 표본크기 n 이 클 때 근사적으로 다음과 같은 분포를 갖는다.

$$Z = \frac{1}{2} \log \left\{ \frac{1+R}{1-R} \right\} \approx AN \left(\frac{1}{2} \log \left\{ \frac{1+\rho}{1-\rho} \right\}, \frac{1}{n-3} \right)$$

따라서 어느 정도 큰 크기의 표본이 임의의 분포에서 취해졌다고 했을 때 표본상관계수 R 의 Z -변환 통계량의 분산은 ρ 와 무관하며, 표본크기가 클수록 정규분포에 근사한다. 그러면 표본상관계수 R 을 Z -변환한 통계량을 이용하여 다음과 같은 EWMA 모형을 고려할 수 있다.

$$EWMA_t = (1 - \lambda) EWMA_{t-1} + \lambda Z_t, \quad t = 1, 2, \dots \quad (3)$$

단, $EWMA_t$: t 시점의 EWMA 값

λ : 평활모수 ($0 < \lambda \leq 1$)

$$R_t = \frac{\sum_{j=1}^n (x_{tj} - \bar{x}_t)(y_{tj} - \bar{y}_t)}{\sqrt{\sum_{j=1}^n (x_{tj} - \bar{x}_t)^2} \sqrt{\sum_{j=1}^n (y_{tj} - \bar{y}_t)^2}}$$

Z_t : t 시점에서 계산된 두 변수 X, Y 의 표본상관 계수 R_t 에 대한

$$\text{변환값 } \frac{1}{2} \log \left\{ \frac{1 + R_t}{1 - R_t} \right\}$$

이와 같이 상관계수의 변동을 감지하기 위해 목표치 ρ 의 값에 따라 표본상관계수 R_t 에 대해 T -변환과 Z -변환한 통계량을 이용한 EWMA 모형을 고려할 수 있을 것이다.

3.2. 상관계수에 대한 EWMA 관리한계

앞 절에서는 표본상관계수를 변환한 통계량의 분포와 이에 대한 EWMA 관리모형을 제시하였다. 따라서 모상관계수 ρ 의 이상증감의 변화를 감지하기 위한 식(2), 식(3)의 EWMA 모형들에 대한 3σ 관리한계를 아래 식에 의해 각각 구하면 될 것이다.

$$\mu_{EWMA} \pm 3\sigma_{EWMA}$$

우선 목표치 $\rho = 0$ 인 경우의 EWMA 모형 식(2)을 고려해 보자. n 이 작은 경우

$$T_t = \frac{\sqrt{n-2}R_t}{\sqrt{1-R_t^2}} \sim t(n-2) \text{이므로}$$

$$\begin{aligned}\mu_{EWMA} &= 0 \\ \sigma_{EWMA}^2 &= \left(\frac{\lambda}{2-\lambda}\right)[1-(1-\lambda)^{2t}]Var\left(\frac{\sqrt{n-2}R_t}{\sqrt{1-R_t^2}}\right) \\ &= \left(\frac{\lambda}{2-\lambda}\right)[1-(1-\lambda)^{2t}]\left(\frac{n-2}{n-4}\right) \\ &\doteq \left(\frac{\lambda}{2-\lambda}\right)\left(\frac{n-2}{n-4}\right), \quad t \rightarrow \infty\end{aligned}$$

와 같다. 따라서, t 가 충분히 클 때는

$$\sigma_{EWMA} = \sqrt{\left(\frac{\lambda}{2-\lambda}\right)\left(\frac{n-2}{n-4}\right)}$$

와 같다. 따라서 n 이 크지 않은 경우 식(2)의 EWMA 관리도의 관리한계는

$$\begin{aligned}CL &= 0 \\ UCL &= 3\sqrt{\left(\frac{\lambda}{2-\lambda}\right)\left(\frac{n-2}{n-4}\right)} \\ LSL &= -3\sqrt{\left(\frac{\lambda}{2-\lambda}\right)\left(\frac{n-2}{n-4}\right)}\end{aligned} \quad (4)$$

와 같다.

다음으로 목표치가 $\rho = \rho_0 \neq 0$ 인 경우, n 이 비교적 큰 경우에 EWMA 모형 식(3)을 고려하면 다음과 같다.

$$Z_t = \frac{1}{2} \log\left\{\frac{1+R_t}{1-R_t}\right\} \approx AN\left(\frac{1}{2} \log\left\{\frac{1+\rho_0}{1-\rho_0}\right\}, \frac{1}{n-3}\right)$$

이므로

$$\begin{aligned}\mu_{EWMA} &= \frac{1}{2} \log\left\{\frac{1+\rho_0}{1-\rho_0}\right\} \\ \sigma_{EWMA}^2 &= \left(\frac{\lambda}{2-\lambda}\right)[1-(1-\lambda)^{2t}]Var\left(\frac{1}{2} \log\left\{\frac{1+R_t}{1-R_t}\right\}\right) \\ &\doteq \left(\frac{\lambda}{2-\lambda}\right)[1-(1-\lambda)^{2t}]\left(\frac{1}{n-3}\right) \\ &\doteq \left(\frac{\lambda}{2-\lambda}\right)\left(\frac{1}{n-3}\right), \quad t \rightarrow \infty\end{aligned}$$

와 같다. 따라서 관리한계는

$$\begin{aligned} CL &= \frac{1}{2} \log \left\{ \frac{1 + \rho_0}{1 - \rho_0} \right\} \\ UCL &= CL + 3 \sqrt{\left(\frac{\lambda}{2 - \lambda} \right) \left(\frac{1}{n - 3} \right)} \\ LSL &= CL - 3 \sqrt{\left(\frac{\lambda}{2 - \lambda} \right) \left(\frac{1}{n - 3} \right)} \end{aligned} \quad (5)$$

와 같이 얻을 수 있다. 따라서 식(4)와 식(5)의 관리한계선에 의한 EWMA 관리도를 그려 상관계수의 이상변동을 감지할 수 있다.

4. 상관계수 변동에 대한 ARL 비교 시뮬레이션

어떤 두 특성변수 X 와 Y 가 각각 주변분포(marginal distribution)에 의해 관리되었을지라도, 이 두 특성변수의 선형성을 관리하고자 할 경우에는 주변분포에 의한 관리만으로는 감지할 수 없으며 공정관리자가 이를 탐지하고 싶다고 하자. 이 경우 두 특성변수들이 각 변수의 평균을 중심으로 선형형태를 보이지 않고 분포하도록 관리하고자 하는 것이다. 따라서 두 특성변수 X, Y 의 선형관계 측도인 상관계수가 0을 중심으로 분포하는가를 조사하는 공정관리를 하여 주변분포에 의한 관리를 보완하고자 하는 것이다. 두 특성변수 X, Y 의 분포가 $BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho_{xy})$ 이라면, 일반성을 잃지 않고 이를 표준화한 분포가 이변량 정규분포 $BN(0, 0, 1, 1, \rho)$ 를 따른다고 가정할 수 있다. 만약 공정관리자의 목표가 공정라인에서 두 특성변수간에 뚜렷한 선형관계가 나타나지 않는 것이라면, 목표치를 $\rho = \rho_0 = 0$ 으로 가정할 수 있다. 이러한 가정 하에서 각 특성변수들의 주변 분포는 서로 독립인 표준정규분포를 따를 것으로 기대되어진다. 따라서 본 시뮬레이션에서는 목표치 $\rho = \rho_0 = 0$ 를 갖는 여러 가지의 부분군 크기에 대해, ρ 의 변동을 증가시킴에 따라 각 평활모수 λ 에 대하여 ARL의 변화를 살펴 보았다. 'Hall & Martin(1989)'은 이변량 정규분포인 경우 비교적 적은 표본크기 일지라도 Z -변환한 표본상관계수 분포는 근사적 정규분포임을 이용하여 소표본에서의 비모수적 볼스트랩 신뢰구간에 대한 시뮬레이션을 하였다. 두 특성변수들을 각 상관계수의 변동값 $\Delta\rho$ 과 평활모수값에 따라 이변량 정규분포 $BN(0, 0, 1, 1, \Delta\rho)$ 에서 각 시점에 대하여 표본크기 $n=5, n=10, n=25, n=50$ 을 생성하였다. 생성된 두 특성변수들의 상관계수값을 각 k 시점에 대하여 관측하여 T -변환과 Z -변환을 한 후 이를 식(2)와 식(3)의 EWMA 모형에 적용하여 값을 구한다. 그리고 식(4)와 식(5)의 관리한계를 벗어날 때까지의 시점에 대하여 $\rho \neq 0$ 이라고 할 수 있을 때까지의 런

길이를 1000번 반복 시행하여 평균값 ARL과 표준편차 SDRL(괄호 속의 값)을 구한 결과가 다음의 표들에 제시되어 있다. 이 때 EWMA 모형의 평활모수 λ 는 $n \geq 50$ 인 예비조사를 통한 반복적 최소오차제곱법에 의해 미리 추정되어 있어야 하는데, 본 논문에서는 평활모수의 추정에 관심을 두지 않으므로 평활모수 λ 값을 0.1~0.5와 Shewhart 관리도와 비교를 위해서 1인 경우까지 고려했다.

$\lambda = 1$ 인 경우의 EWMA값은 변환한 표본상관계수값들은 Shewhart 값과 같으므로, 표에서 살필 수 있듯이 상관계수의 변동이 없을 때(0%일 때)에는 표본크기에 관계없이 EWMA 관리도의 ARL이 Shewhart 관리도의 ARL보다 큰 것을 관측할 수 있다. 따라서 EWMA 관리도가 Shewhart 관리도보다 오경보(false alarm)를 덜 준다고 할 수 있으므로 수행도가 좋음을 알 수 있다.

$n = 5$ 인 경우인 <표 1>과 <표 5>로부터 상관계수의 변동(0.02~0.32)에 대하여 EWMA 관리도의 ARL이 Shewhart 관리도의 ARL보다 길어 공정이상 상태를 늦게 탐지한다.

$n = 10$ 인 경우인 <표 2>와 <표 6>에서 상관계수의 변동값이 0.08이하인 경우에는 Shewhart 관리도의 수행도가 더 좋으며 0.16이상이고 $\lambda = 0.1$ 이 아닌 경우에는 EWMA 관리도의 ARL이 Shewhart 관리도의 ARL보다 더 작게 관측되어 EWMA 관리도가 민감하게 이상을 감지하여 수행도가 더 좋음을 알 수 있다. 특히, Z-변환은 0.16이상의 모든 λ 에 대하여 EWMA 관리도의 ARL이 더 작았다.

< 표 1 > $n = 5, \rho_0 = 0, T$ -변환

ρ 의 변동값 ($\Delta\rho$)	λ					
	0.1	0.2	0.3	0.4	0.5	1
0	506.541 (489.433)	234.91 (220.053)	149.713 (141.053)	116.459 (111.787)	104.019 (104.522)	76.888 (78.849)
0.02	471.273 (469.545)	223.512 (223.794)	152.316 (148.089)	128.376 (132.274)	102.805 (99.939)	76.681 (76.098)
0.04	445.710 (414.485)	217.632 (206.154)	145.719 (153.634)	117.906 (112.815)	102.361 (100.161)	73.853 (74.517)
0.08	393.659 (384.587)	209.59 (210.398)	146.748 (146.818)	111.760 (114.261)	97.842 (95.485)	73.705 (72.657)
0.16	245.156 (249.267)	136.121 (138.825)	109.405 (108.024)	93.447 (89.497)	83.069 (82.343)	65.43 (63.896)
0.32	68.964 (68.175)	53.231 (52.374)	51.820 (50.659)	47.561 (48.058)	49.520 (47.949)	43.239 (43.704)
0.50	18.652 (14.720)	17.246 (14.664)	18.315 (16.400)	19.662 (19.305)	20.668 (19.739)	24.211 (24.082)
0.64	9.239 (4.817)	8.883 (6.206)	8.811 (7.278)	9.390 (7.727)	10.570 (9.605)	13.647 (12.614)
0.80	5.093 (2.035)	4.437 (2.285)	4.201 (2.418)	4.130 (2.679)	4.332 (3.019)	5.869 (5.409)

< 표 2 > $n=10, \rho_0=0, T$ -변환

ρ 의 변동값 ($\Delta\rho$)	λ					
	0.1	0.2	0.3	0.4	0.5	1
0	1040.78 (1045.83)	445.627 (437.231)	273.185 (266.673)	218.634 (219.288)	175.429 (176.287)	125.642 (126.604)
0.02	988.923 (967.702)	400.163 (374.833)	264.125 (269.103)	211.392 (214.272)	172.071 (168.619)	124.447 (123.795)
0.04	720.149 (690.013)	353.465 (337.758)	241.11 (248.279)	189.25 (183.335)	160.719 (158.638)	127.053 (123.824)
0.08	344.168 (331.965)	199.094 (206.487)	172.951 (166.678)	141.231 (138.842)	132.78 (122.426)	109.497 (115.542)
0.16	82.326 (79.307)	65.99 (62.388)	67.464 (66.124)	69.118 (63.919)	65.702 (65.065)	72.602 (71.661)
0.32	13.648 (9.582)	13.184 (10.108)	13.489 (11.353)	14.541 (12.608)	17.071 (15.648)	29.204 (29.025)
0.50	6.008 (2.289)	5.077 (2.464)	4.931 (2.770)	5.048 (3.417)	5.395 (3.914)	9.434 (8.974)
0.64	4.021 (1.341)	3.246 (1.355)	2.975 (1.351)	2.848 (1.439)	2.866 (1.712)	3.801 (3.202)
0.80	2.595 (0.796)	2.034 (0.697)	1.798 (0.701)	1.672 (0.701)	1.601 (0.709)	1.567 (0.926)

< 표 3 > $n=25, \rho_0=0, T$ -변환

ρ 의 변동값 ($\Delta\rho$)	λ					
	0.1	0.2	0.3	0.4	0.5	1
0	1379.980 (1313.34)	639.601 (599.422)	443.833 (441.134)	365.815 (380.307)	314.229 (304.121)	222.743 (208.436)
0.02	996.595 (1024.44)	542.589 (499.255)	354.154 (346.388)	317.997 (320.048)	279.166 (285.872)	227.844 (228.386)
0.04	481.992 (505.389)	299.851 (279.248)	248.728 (243.113)	217.345 (221.584)	210.03 (205.653)	195.574 (190.310)
0.08	118.896 (109.171)	107.653 (111.005)	96.461 (88.448)	106.487 (108.962)	108.089 (105.979)	142.869 (145.136)
0.16	21.193 (16.978)	20.386 (17.469)	22.384 (19.921)	25.819 (23.492)	29.584 (27.653)	60.228 (56.903)
0.32	6.175 (2.278)	5.061 (2.138)	4.96 (2.485)	5.012 (3.117)	5.376 (3.891)	10.295 (9.755)
0.50	3.302 (0.875)	2.606 (0.836)	2.364 (0.841)	2.215 (0.890)	2.139 (0.972)	2.597 (2.076)
0.64	2.354 (0.570)	1.879 (0.548)	1.651 (0.559)	1.502 (0.579)	1.419 (0.536)	1.308 (0.603)
0.80	1.662 (0.482)	1.211 (0.408)	1.071 (0.257)	1.029 (0.168)	1.025 (0.156)	1.008 (0.089)

< 표 4 > $n=50, \rho_0=0, T$ -변환

ρ 의 변동값 ($\Delta\rho$)	λ					
	0.1	0.2	0.3	0.4	0.5	1
0	1478.12 (1502.61)	687.63 (668.2)	490.143 (519.653)	440.846 (438.83)	358.967 (347.547)	286.095 (276.018)
0.02	720.105 (735.632)	460.335 (448.602)	357.94 (370.234)	318.159 (318.644)	299.709 (297.007)	258.619 (248.442)
0.04	249.123 (250.523)	194.928 (194.917)	186.799 (175.077)	169.012 (159.009)	192.507 (185.429)	207.326 (206.669)
0.08	48.122 (44.791)	43.839 (41.303)	51.507 (49.411)	54.717 (53.263)	66.345 (68.977)	119.778 (127.112)
0.16	9.937 (4.875)	9.172 (5.532)	9.704 (7.267)	10.972 (9.309)	12.060 (10.789)	29.472 (29.482)
0.32	3.911 (1.103)	3.191 (1.071)	2.904 (1.131)	2.725 (1.144)	2.751 (1.351)	3.906 (3.505)
0.50	2.339 (0.506)	1.887 (0.480)	1.607 (0.563)	1.443 (0.547)	1.354 (0.517)	1.243 (0.544)
0.64	1.82 (0.402)	1.268 (0.445)	1.097 (0.296)	1.052 (0.222)	1.025 (0.156)	1.019 (0.137)
0.80	1.079 (0.270)	1.001 (0.032)	1 (0)	1 (0)	1 (0)	1 (0)

< 표 5 > $n=5, \rho_0=0, Z$ -변환

ρ 의 변동값 ($\Delta\rho$)	λ					
	0.1	0.2	0.3	0.4	0.5	1
0	1614.94 (1610.09)	695.683 (726.899)	442.624 (439.063)	338.187 (325.525)	267.918 (270.324)	185.257 (187.637)
0.02	1616.98 (1564.67)	657.57 (676.712)	450.664 (432.779)	313.166 (305.302)	257.428 (254.969)	182.84 (188.066)
0.04	1389.12 (1332.78)	591.334 (617.974)	419.867 (319.125)	342.933 (344.642)	270.393 (289.313)	188.187 (187.62)
0.08	864.238 (852.956)	470.669 (464.326)	354.138 (331.517)	291.026 (299.386)	235.659 (238.378)	183.061 (175.622)
0.16	309.807 (302.359)	223.53 (214.989)	194.258 (182.465)	179.575 (179.377)	166.905 (167.441)	161.598 (157.145)
0.32	53.069 (48.57)	51.079 (48.017)	59.921 (57.351)	70.66 (69.977)	75.923 (70.855)	105.527 (107.004)
0.50	14.758 (9.782)	14.722 (11.206)	17.520 (15.224)	19.551 (18.027)	25.694 (24.733)	57.327 (62.460)
0.64	8.480 (3.334)	7.469 (3.885)	8.044 (5.366)	9.263 (7.311)	11.132 (9.814)	32.703 (34.043)
0.80	5.408 (1.539)	4.405 (1.617)	4.141 (1.798)	4.266 (2.303)	4.721 (3.003)	12.087 (11.399)

< 표 6 > $n=10, \rho_0=0, Z$ -변환

ρ 의 변동값 ($\Delta\rho$)	λ					
	0.1	0.2	0.3	0.4	0.5	1
0	1523.26 (1601.65)	681.164 (648.902)	461.641 (433.697)	375.647 (372.758)	327.137 (341.672)	259.908 (278.133)
0.02	1268.98 (1261.07)	589.695 (584.881)	445.265 (456.957)	354.762 (359.010)	318.148 (321.673)	244.752 (234.043)
0.04	939.253 (951.167)	487.896 (498.873)	377.276 (372.868)	338.308 (334.655)	300.992 (294.127)	234.934 (236.855)
0.08	386.145 (382.147)	265.699 (270.348)	241.761 (247.693)	225.014 (224.653)	212.736 (223.734)	214.577 (215.309)
0.16	85.585 (79.851)	77.859 (73.156)	79.912 (80.701)	92.258 (88.689)	94.763 (95.573)	151.034 (151.476)
0.32	13.698 (8.822)	13.43 (10.602)	15.189 (13.267)	16.841 (15.068)	21.641 (19.976)	47.958 (49.265)
0.50	6.244 (2.130)	5.296 (2.218)	5.310 (2.682)	5.470 (3.510)	5.936 (3.865)	14.908 (14.946)
0.64	4.401 (1.186)	3.502 (1.186)	3.202 (1.268)	3.069 (1.365)	3.096 (1.545)	5.465 (5.026)
0.80	3.022 (0.659)	2.368 (0.591)	2.134 (0.617)	1.925 (0.634)	1.842 (0.753)	1.908 (1.245)

< 표 7 > $n=25, \rho_0=0, Z$ -변환

ρ 의 변동값 ($\Delta\rho$)	λ					
	0.1	0.2	0.3	0.4	0.5	1
0	1583.2 (1707.87)	709.208 (693.149)	509.752 (497.670)	428.524 (414.773)	427.1 (418.799)	325.895 (312.731)
0.02	1021.01 (1060.49)	523.907 (543.770)	436.232 (458.168)	376.674 (381.705)	367.367 (347.074)	309.706 (312.888)
0.04	493.278 (500.673)	325.390 (344.328)	300.608 (293.490)	284.319 (274.712)	267.532 (256.591)	269.349 (253.582)
0.08	128.107 (122.201)	111.132 (104.121)	110.992 (108.493)	112.459 (106.019)	130.533 (130.202)	198.522 (194.343)
0.16	20.470 (15.900)	20.643 (18.273)	22.9 (20.725)	26.993 (26.042)	33.142 (32.845)	78.171 (83.307)
0.32	6.104 (2.031)	5.36 (2.318)	4.86 (2.467)	4.976 (2.908)	5.21 (3.608)	14.063 (13.568)
0.50	3.458 (0.826)	2.785 (0.797)	2.515 (0.853)	2.307 (0.845)	2.219 (0.978)	2.755 (2.193)
0.64	2.500 (0.559)	2.024 (0.449)	1.776 (0.519)	1.573 (0.586)	1.508 (0.573)	1.384 (0.713)
0.80	1.966 (0.234)	1.388 (0.487)	1.156 (0.363)	1.065 (0.247)	1.029 (0.168)	1.016 (0.133)

< 표 8 > $n=50$, $\rho_0=0$, Z -변환

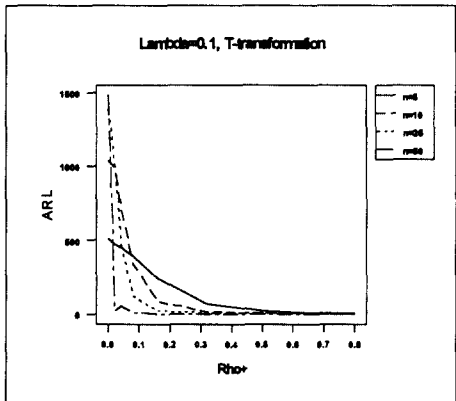
ρ 의 변동값 ($\Delta\rho$)	λ					
	0.1	0.2	0.3	0.4	0.5	1
0	1551.84 (1527.8)	720.472 (709.742)	543.063 (545.454)	451.586 (454.992)	427.143 (436.703)	366.677 (385.657)
0.02	796.796 (747.683)	453.284 (444.766)	374.517 (358.091)	343.461 (344.168)	342.404 (340.943)	332.321 (351.138)
0.04	251.605 (253.468)	206.938 (196.831)	208.183 (212.702)	188.179 (189.182)	202.154 (211.685)	262.406 (264.766)
0.08	45.955 (41.635)	45.415 (40.457)	49.416 (48.103)	60.524 (58.058)	71.197 (71.765)	131.992 (139.806)
0.16	9.924 (4.903)	8.98 (5.535)	9.554 (7.017)	10.55 (8.354)	12.762 (10.641)	34.768 (35.590)
0.32	3.991 (1.047)	3.217 (1.041)	2.911 (1.081)	2.756 (1.212)	2.68 (1.357)	4.359 (3.889)
0.50	2.403 (0.541)	1.950 (0.460)	1.715 (0.544)	1.488 (0.535)	1.396 (0.532)	1.273 (0.575)
0.64	1.955 (0.251)	1.404 (0.493)	1.116 (0.320)	1.074 (0.262)	1.04 (0.196)	1.017 (0.129)
0.80	1.253 (0.435)	1.005 (0.070)	1.003 (0.055)	1 (0)	1 (0)	1 (0)

마찬가지로 $n=25$ 일 때는 상관계수 변동값이 0.04이하인 경우와 $n=50$ 일 때 0.02이하인 경우는 Shewhart 관리도의 수행도가 더 좋으며, $n=25$ 일 때는 0.08이상과 $n=50$ 일 때 0.04이상인 경우는 EWMA 관리도가 수행도가 보다 좋음을 <표 3>, <표 7> 그리고 <표 4>, <표 8>에서 알 수 있다. 따라서 표본크기를 적당히 크게(여기서는 $n \geq 10$) 해 주면 EWMA 관리도가 상관계수의 0.06이상의 변동을 관리하기에는 더욱 좋다고 할 수 있다. 상관계수의 변동율이 증가함에 따라 당연히 ARL은 감소를 해야한다. 그런데 각 경우의 표에서 ARL이 감소하는 추세에서 간혹 증가를 하는 경우가 있다. 이는 표본크기가 적을수록, T -변환일수록, 상관계수의 변동값이 작을수록, λ 값이 클 때 빈번하게 관찰되어진다. 이런 경우에 변동이 증가된 경우의 ARL값이 비록 이전의 변동의 ARL에 비하여 증가는 되었지만, 1000개에 대한 RL(run length)의 표준편차 값은 큰 경우가 대부분이었다. 이러한 결과는 표본의 크기가 작을 때와 상관계수의 변동 증가값이 비교적 작으므로, 시뮬레이션의 난수발생시에 작은 표본크기로 인한 결과로 생각되어지는데, 이는 $n=50$ 일 때는 한 번도 관측되지 않았으므로 표본크기가 클수록 그 관측횟수가 적어짐으로써 판단될 수 있다. 그리고 ARL값의 Z -변환 시에는 $n=5$ 인 경우를 제외하고는 한 번도 관측되지 않으므로 T -변환일 때와 λ 값이 클 때 불안정함을 알 수 있다.

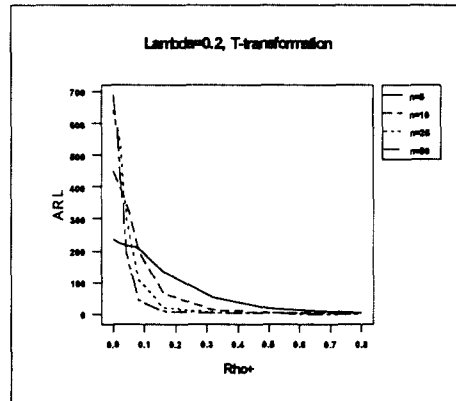
한편, 이변량 정규분포를 따르는 변수들간의 상관계수의 변동에 따른 ARL들이 각

λ 값과 표본크기에 대해 T -변환과 Z -변환을 사용한 경우에 어떤 차이를 보이는가를 쉽게 파악하기 위한 플랏이 <그림 1>~<그림 12>에 제시되어 있다. 6가지의 평활모수 λ 값에 대한 T -변환 홀수번호 <그림 1>~<그림 11>과 Z -변환 짝수번호 <그림 2>~<그림 12>를 보자. 모든 λ 값에 대하여, 표본크기가 $n=5$ 인 경우 T -변환 ARL은 거의 변화를 보이지 않고 일정하여 상관계수의 변동을 탐지에 민감하지 못함을 알 수 있다. 반면 Z -변환의 ARL은 표본크기가 작은 경우에도 비교적 안정된 곡선을 나타내어 좋은 수행도를 시사하고 있다. 이는 표본크기가 $n=10$ 이상인 경우에도 비슷한 양상을 띄며, 표본크기가 클수록 더욱 안정된 곡선으로 올바른 탐지도가 높음을 알 수 있다. 그러나 λ 값이 비교적 큰 경우인 0.5~1의 경우에는 $n=25$ 인 경우의 두 변환 모두 ARL이 상관계수의 변동에 둔감하게 변하는 일정한 양상을 보이고 있다. 따라서 Shewhart 관리도는 표본크기가 상당히 클 때(여기서는 $n=50$ 일 때)만 좋은 수행도를 보인다고 할 수 있다. 그리고 λ 값이 0.1~0.4일 때, T -변환은 $n=25$ 이상일 때 안정된 ARL을 갖지만, Z -변환은 $n=10$ 일 때도 비교적 바람직한 ARL을 나타내어 보다 민감하게 상관계수의 변동을 탐지할 수 있음을 알 수 있다.

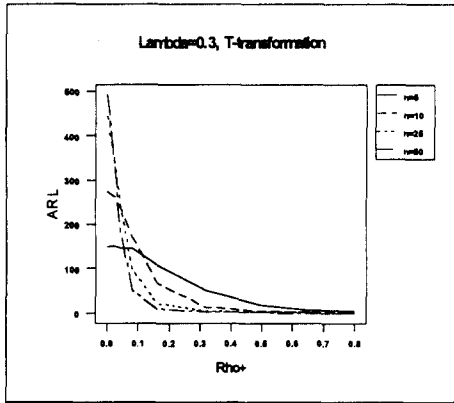
이와 같이 ARL관점에서 볼 때 Z -변환이 더욱 안정적이고 탐지도가 높았다. 이변량 정규분포를 따르는 변수들간의 표본상관계수에 대한 Z -변환은 상당히 작은 수의 표본일지라도 근사적 정규분포를 따르므로, 분포의 가정 하에서 표본크기가 너무 작지 않은 경우에는 Z -변환을 한 표본상관계수가 상관계수값의 변동을 감지하는데 이용될 수 있을 것이다. 'Hall & Martin(1989)'의 시뮬레이션 연구를 참고하면 대략 이변량 정규분포하에서 추출된 표본크기 $n=10$ 이상의 표본상관계수값을 사용하여 상관계수의 변동을 관리할 수 있을 것이다. 또한 Z -변환은 T -변환에 비해 다양한 목표치의 상관계수값을 관리할 수 있어 응용의 폭이 크고, 변환계산에 드는 메모리와 CPU 사용시간도 거의 차이를 보이지 않는다.



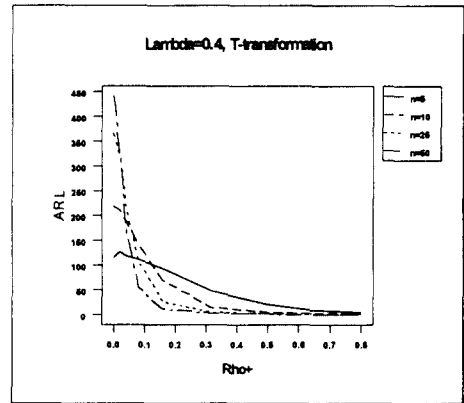
< 그림 1 > $\lambda=0.1$, T 통계량



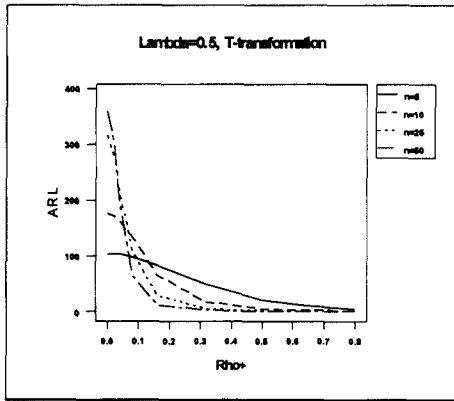
< 그림 2 > $\lambda=0.2$, T 통계량



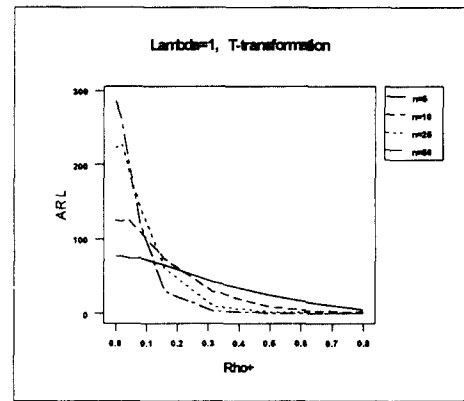
< 그림 3 > $\lambda=0.3$, T 통계량



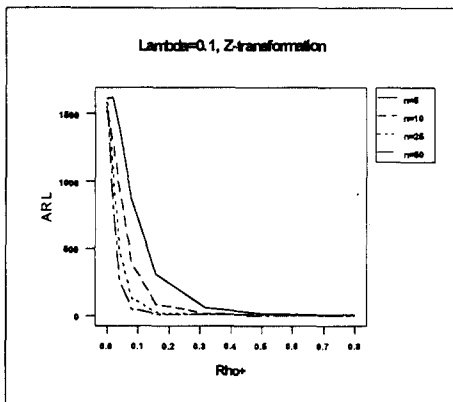
< 그림 4 > $\lambda=0.4$, T 통계량



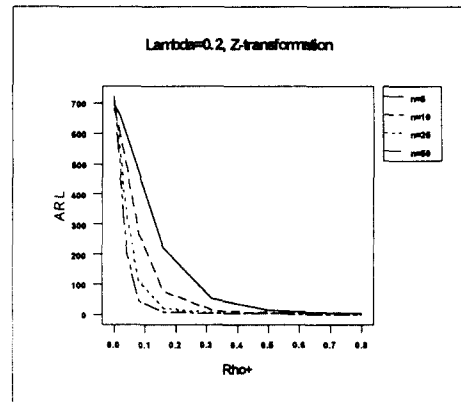
< 그림 5 > $\lambda=0.5$, T 통계량



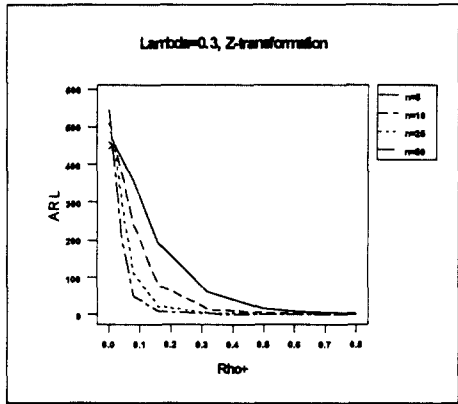
< 그림 6 > $\lambda=1$, T 통계량



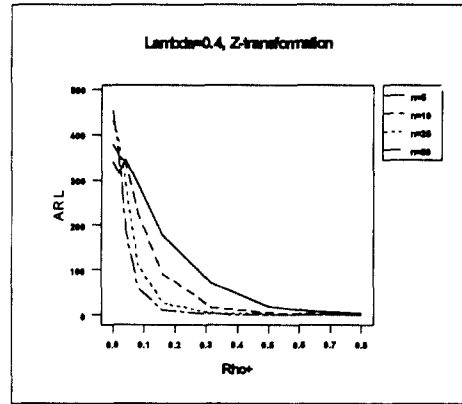
< 그림 7 > $\lambda=0.1$, Z 통계량



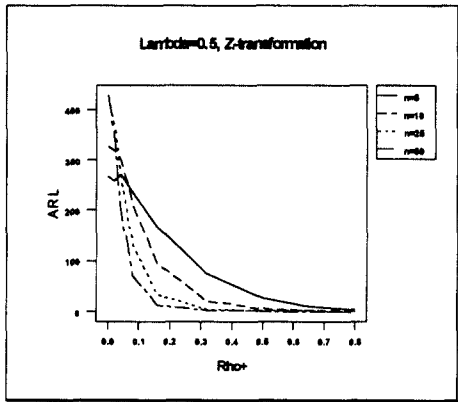
< 그림 8 > $\lambda=0.2$, Z 통계량



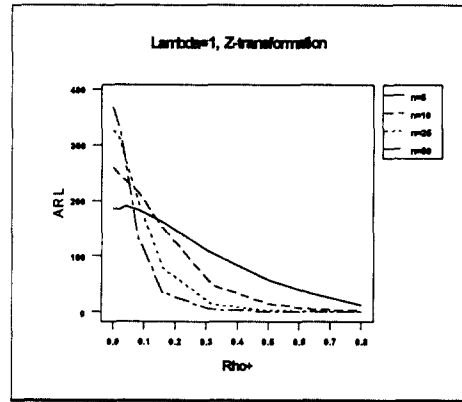
< 그림 9 > $\lambda=0.3$, Z통계량



< 그림 10 > $\lambda=0.4$, Z통계량



< 그림 11 > $\lambda=0.5$, Z통계량



< 그림 12 > $\lambda=1$, Z통계량

5. 결론

본 논문에서는 공정관리자가 상관계수의 변동을 관리하고자 할 때의 EWMA 모형에 대하여 연구하였다. 상관계수값의 목표치가 $\rho = \rho_0 = 0$ 일 때 표본상관계수 R_t 를 기초로 T-변환과 Z-변환을 기초로 EWMA 관리모형에 적용하여 그 수행도를 비교하였다. 이변량 표준정규분포를 따르는 변량에서 추출한 적당한 부분군의 크기 하에서 상관계수의 변동이 다소 클 때에는 EWMA 관리도의 ARL이 Shewhart 관리도의 ARL보다 짧게 나오므로, 공정이 이상 상태에 놓여 있을 때 보다 민감하게 이를 탐지한다. 따라서 상관계수의 변화가 다소 큰 경우의 공정관리에서는 Shewhart 관리도보다 EWMA 관리도가 더욱 유용할 것이다. 그리고 상관계수의 변동이 없는 경우에도

Z-변환을 취한 ARL이 크게는 T-변환의 ARL보다 2~3배로, Z-변환을 쓴 경우가 훨씬 적은 수의 오경보를 주어 수행도가 더 좋음을 알릴 수 있다. 그리고 상관계수의 변동이 큰 경우에는 부분군의 크기가 비교적 작은 경우의 ARL값은 T-변환을 사용한 경우가 Z-변환의 경우보다 약간 더 작음을 알릴 수 있으므로, T-변환 EWMA 관리모형을 사용하는 것이 더 빨리 공정의 이상을 탐지하여 좋은 수행도를 갖지만 변동에 민감하지는 않다. 그리고 부분군의 크기가 크면 두 변환에 대한 ARL은 큰 차이를 보이지 않는다. 하지만 n 이 작은 경우 상관계수의 변동이 다소 커야 Shewhart 관리도의 수행도보다 좋아지고, n 이 큰 경우는 비교적 작은 상관계수의 변동도 민감하게 탐지할 수 있다. 따라서 상관계수의 변화가 다소 큰 경우만 관리하고자 할 때는 부분군의 크기가 작은 T-변환을 쓰는 것이 더 빨리 변동이상을 탐지할 것이고, 작은 변동을 관리해야 할 경우와 같이 민감한 공정관리 시에는 Z-변환을 써야 할 것이다.

결론적으로 여러 가지 표본크기를 갖고 시뮬레이션을 실시한 결과, 상관계수의 변화가 적은 경우 EWMA 관리도의 ARL은 Shewhart 관리도의 ARL보다 더 큰 값을 갖으므로 상관계수의 미세한 변동이 발생했을 때 Shewhart 관리도에 비해 적은 오경보(false alarm)를 주는 것을 알 수 있었다. 그리고 상관계수의 변동이 큰 경우의 EWMA 관리도의 ARL은 Shewhart의 ARL보다 작은 값을 갖는데, 이것은 상관계수의 변동이 다소 크게 생겼을 때도 EWMA 관리도가 Shewhart 관리도보다 더 민감하게 공정이상을 탐지하여 높은 수행도를 갖음을 확인할 수 있었다. 그리고 상관계수의 이상변화 탐지를 위한 두 변환법 중 Z-변환이 T-변환보다 더욱 안정적으로 민감하게 변동이상을 탐지함을 알 수 있었다.

이와 같이 본 논문에서는 상관계수의 변동의 관리를 위한 EWMA 모형을 제안하고, 실제 현장에서 사용되기 위해 EWMA 관리모형의 ARL을 시뮬레이션 결과로써 구하여 보았다. 따라서 공정관리자들은 각 시점에 대해 부분군 크기 n 의 표본을 얻어 T-변환, Z-변환을 한 후 MINITAB, SAS 등과 같은 통계팩키지로 간단히 온라인(on-line)공정관리를 실시할 수 있을 것이다. 본 논문에서는 표본들이 이변량 정규분포를 따르며 목표치가 $\rho = \rho_0 = 0$ 인 경우의 상관계수 증가 변동문제를 중심으로 연구하였으나, 목표치가 $\rho = \rho_0 \neq 0$ 인 경우나 상관계수가 감소하는 경우에도 어렵지 않게 적용될 수 있을 것이다.

참고문헌

- [1] 장덕준(1994), "가변추출간격을 이용한 다변량 관리도에 관한 연구," 「서울대학교 박사학위논문」.
- [2] Crowder, S.V.(1987), "A Simple Method for Studying Run Length Distributions of Exponentially Weighted Moving Average Charts," *Technometrics* 29,

- pp. 393~407.
- [3] Crowder, S.V.(1989), "Design of Exponentially Weighted Moving Average Schemes," *Journal of Quality Technology* 21, pp. 155~162.
 - [4] Crowder, S.V. and Hamilton, M.D.(1992), "An EWMA for Monitoring a Process Standard Deviation," *Journal of Quality Technology* 24, pp. 12~21.
 - [5] Fisher, R.A.(1915), "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, Vol. 10, pp. 507~521.
 - [6] Fisher, R.A.(1921), "On the probable error of a coefficient of correlation deduced from a small sample," *Metron*, Vol. 1, pp. 3~32.
 - [7] Hall, P. and Martin, M.A.(1989), "Better Nonparametric Bootstrap Confidence Intervals for the Correlation Coefficient," *Journal of Statistical and Computational Simulation*, Vol. 33, pp. 161~172.
 - [8] Hunter J.S.(1986), "The Exponentially Weighted Moving Average," *Journal of Quality Technology*, 18, pp. 203~210.
 - [9] Lucas, J.M. and Saccucci, M.S.(1990), "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements." *Technometrics* 32, pp. 1~12.
 - [10] Montgomery, D.C. and Mastrangelo, C.M.(1991), "Some Statistical Process Control Methods for Autocorrelated data," *Journal of Quality Technology*, Vol. 23, No. 3, pp. 179~204.
 - [11] Roberts, S.W.(1959), "Control Chart Tests Based on Geometric Moving Averages." *Technometrics* 1, pp. 239~250.
 - [12] Robinson, P.B. and Ho, T.Y.(1978), "Average Run Lengths of Geometric Moving Average Charts by Numerical Methods." *Technometrics* 20, pp. 85~93.
 - [13] *SAS/IML User's Guide*(1988), Release 6.03 Edition, SAS Institute Inc.
 - [14] *SAS/QC Software Reference*(1989), Version 6 First Edition, SAS Institute Inc.