The Korean Communications
in Statistics Vol. 10 No. 2, 2003
pp. 497-505

Influence Function on Tolerance Limit

Honggie Kim[1], Yun Hee Lee[2], Hee Sung Shin[3], Sounki Lee[4]

Abstract

Under normality assumption, the tolerance interval for a future observation is sometimes of great interest in statistics. In this paper, we state the influence function on the standard deviation σ, and use it to derive the influence function on tolerance limits. Simulation study shows that the two influence functions perform very well.

Keywords : Influence function, tolerance interval, tolerance limit, confidence interval

1. 서론

정규분포의 평균에 대한 신뢰구간과 정규분포에서의 허용구간은 대부분의 통계서적에서 주로 다루어지고 있고, 그 중에서 허용구간은 통계적 품질공정 또는 실험분야에서 매우 중요시된다. 허용구간은 표본으로부터 미래 관찰치가 존재하는 구간을 계산한다. 이 허용구간에 임의의 관찰치가 미치는 영향을 측정하기 위해 영향함수를 이용할 수 있다. 각 관찰치에서의 영향함수값은 그 관찰치가 허용구간에 미치는 고유한 영향값으로 정의되며, 이 절대값은 정규분포의 양쪽 꼬리부분에 위치하는 데이터일수록 큰 값을 보인다. 또한 표본 중에서 다른 관찰치들과 편차가 큰 관찰치에서도 영향함수의 값이 매우 극단적인 값으로 나타날 것이므로 그 관찰치는 이상치일 가능성이 높다.

Cook과 Weisberg(1980,1982)는 선형회귀모형에서 회귀계수 벡터의 신뢰영역에 대한 관찰치의 영향을 개체 제거법(case deletion method)을 이용한 신뢰영역의 체적(volume)의 비로 측정하였고, 이 방법은 Ross(1990)에 의해 비선형회귀모형에 활용되었다. 또한 Thomas(1990)는 개체 가중법(case weighting method)을 이용해 동일한 방법으로 일반화 선형모형에 적용하였다. 그 외 Kim(1992)은 대응 분석에서의 영향함수를 유도하였고, Kim과 Lee(1996) 그리고 Kim(1998)은 χ^2 통계량에 대한 영향함수를 유도하였다.

본 논문에서는 Hampel(1974)이 제안한 영향함수의 정의를 이용하여 허용구간의 길이에 관련된 σ와 허용한계에 대한 영향함수를 유도하였고, 모의실험을 통해 정규분포의 확률표본을 대상으로

1) Professor, Department of Statistics, Chungnam National University, Taejon, 360-764, Korea. E-mail: hgkim@stat.cnu.ac.kr
2) Candidate for Ph.D, Department of Statistics, Chungnam National University, Taejon 360-764, Korea.
3) Principle Researcher, Spent Fuel Management Technology Development Division, Korea Atomic Energy Research Institute, Taejon, 305-353, Korea.
4) Candidate for Ph.D, Department of Statistics, Chungnam National University, Taejon, 360-764, Korea.

유도된 영향함수들의 수행능력을 평가하였다.

2. 영향함수

2.1 영향함수 정의

분포함수 F의 함수로 이루어지며 실수공간에서 정의되는 함수를 범함수(functional)라고 한다 (Cook & Weisberg, 1982). 영향함수는 하나의 관찰치가 분포함수 F에서 정의되는 범함수 $T(F)$에 대하여 어느 정도의 영향을 갖는지를 측정한다. 분포함수 F에 임의의 관찰치 x를 추가하면 다음 식과 같이 섭동(pertubation)된 분포함수가 된다.

$$F_\varepsilon = (1-\varepsilon)F + \varepsilon\delta_x \tag{2.1}$$

식(2.1)에서 $0<\varepsilon<1$ 이고, δ_x는 x에서만 확률 1을 갖는 퇴화분포(degenerate distribution)함수이다. 따라서 식(2.1)과 같은 섭동된 분포함수 F_ε에서 범함수 $T(F)$는 섭동된 범함수 $T(F_\varepsilon)$로 정의할 수 있다. Hampel(1974)은 범함수 $T(F)$에 대해 관찰치 x가 갖는 영향을 측정할 수 있는 도구로써 다음 식을 정의하였다.

$$\begin{aligned} IF(x, T) &= \lim_{\varepsilon\to 0}\frac{T(F_\varepsilon)-T(F)}{\varepsilon} \\ &= \lim_{\varepsilon\to 0}\frac{T[(1-\varepsilon)F+\varepsilon\delta_x]-T(F)}{\varepsilon} \end{aligned} \tag{2.2}$$

식(2.2)에서 $IF(x, T)$는 $T(F)$의 일차미분계수에 해당된다. 즉, 추가된 x에 의한 섭동된 범함수 $T(F_\varepsilon)$은 원래의 범함수 $T(F)$에 비해 어느 정도 차이가 있는지를 나타내므로, 이 크기는 $T(F)$에 대한 x의 고유한 영향을 나타낸다. 식(2.2)는 다음 식과 같이 표현할 수 있다.

$$IF(x, T) = \left[\frac{\partial T(F_\varepsilon)}{\partial\varepsilon}\right]_{\varepsilon=0} \tag{2.3}$$

모집단의 평균과 분산은 범함수 $T(F)$에 대응되므로 각각을 범함수 $\mu(F)$와 $\sigma^2(F)$로 표현하면, 평균과 분산에 대한 영향함수는 식(2.2) 또는 식(2.3)을 이용하여 다음 식과 같이 유도할 수 있다[Hampel,1974].

$$IF(x, \mu) = x-\mu \tag{2.4}$$

$$\begin{aligned} IF(x, \sigma^2) &= (x-\mu)^2-\sigma^2 \\ &= [IF(x,\mu)]^2-\sigma^2 \end{aligned} \tag{2.5}$$

식(2.4)에서 볼 수 있듯이 평균에 대한 영향함수는 관찰치 x가 평균 μ와 차이가 클수록 평균에 대해 큰 영향을 갖는 것으로 알 수 있다. 또한 분산에 대한 영향함수도 x에서의 편차제곱 $(x-\mu)^2$과 원래 분산 σ^2의 차이가 클수록 그 관찰치가 분산에 주는 영향이 크다는 것을 알 수 있고, 이것은 식(2.5)와 같이 평균에 대한 영향함수로 표현된다. 평균과 분산에 대한 영향함수는 관찰치에

서의 각 범함수의 편의(bias)로 설명할 수 있다.

2.2 영향함수 확장

가. 표준편차의 영향함수

허용구간의 길이와 허용한계에 대한 영향함수를 구하기 위해서는 표준편차 σ에 대한 영향함수를 알아야 한다. 범함수의 함수는 역시 범함수이다(Kim,1996,1998). 따라서 $\sigma=\sqrt{\sigma^2}$이므로 표준편차 σ는 범함수 $\sigma^2(F)$의 함수이기 때문에 범함수에 해당되어 $\sigma(F)=\sqrt{\sigma^2(F)}$와 같이 표현할 수 있다. 범함수 $\sigma(F)$에 대한 영향함수를 구하기 위하여 섭동된 범함수 $\sigma(F_\varepsilon)$가 필요하고 이것은 $\sqrt{\sigma^2(F_\varepsilon)}$로 추정된다. 따라서 식(2.3)을 이용하여 $\sigma(F)$에 대한 영향함수를 유도하면 다음 식과 같다.

$$\begin{aligned} IF(x,\sigma) &= \left[\frac{\partial\sqrt{\sigma^2(F_\varepsilon)}}{\partial\varepsilon}\right]_{\varepsilon=0} \\ &= \left[\frac{1}{2}\frac{1}{\sqrt{\sigma^2(F_\varepsilon)}}\frac{\partial\sigma^2(F_\varepsilon)}{\partial\varepsilon}\right]_{\varepsilon=0} \\ &= \frac{1}{2\sigma(F)}\cdot IF(x,\sigma^2) \end{aligned} \tag{2.6}$$

식(2.6)에서 볼 수 있듯이 $\sigma(F)$에 대한 영향함수는 $\sigma^2(F)$의 영향함수로 표현된다. 여기서 표준편차에 대한 영향함수는 평균이나 분산에 대한 영향함수와 같이 편의가 아님을 알 수 있다.

나. 허용한계의 영향함수

정규분포의 모수들이 모두 기지이면, 정규확률밀도함수 중심부분의 면적이 α인 구간의 한계는 정확히 $\mu\pm z_{\alpha/2}\sigma$가 된다. 그러나 일반적으로 모수들은 미지이고 이것은 표본으로부터 추정되므로, 추정된 구간의 한계 $\bar{x}\pm z_{\alpha/2}s$에서의 면적은 정확히 α가 아닐 것이다. 따라서 표본으로부터 정규모집단의 중심 면적이 적어도 α가 되도록 γ의 신뢰도로 추정한 구간을 허용구간 $(\bar{x}-k_{\alpha/\gamma,n}s,\ \bar{x}+k_{\alpha/\gamma,n}s)$ 이라 하고, 구간의 한계를 허용한계라고 한다. 여기서 $k_{\alpha/\gamma,n}$는 허용계수이고 n은 표본크기를 나타낸다. 허용한계는 정확히 α 면적을 만족하는 경계값인 $\mu\pm z_{\alpha/2}\sigma$를 γ의 신뢰도로 추정한 것이라고 할 수 있다. 양측 허용구간과 단측 허용구간의 허용계수 $k_{\alpha/\gamma,n}$는 각각 Wald 및 Wolfwiz(1946)와 Guenther(1992)에 의해 정의되었다.

단측 허용한계 $\bar{x}+k_{\alpha/\gamma,n}s$ 에 대한 영향함수를 구하기 위하여 우선 정규분포에서 α 분위수(quantile)인 $\mu+z_\alpha\sigma$ 에 대한 영향함수를 구해보자. $\mu+z_\alpha\sigma$ 는 범함수 $\mu(F)$와 $\sigma(F)$의 함수이므로 범함수에 해당되고 특히 선형결합(linear combination) 형태이다. $\mu+z_\alpha\sigma$ 를 $T(F)$라고 하면, 이것의 섭동된 분포함수 F_ε에서의 범함수는 다음 식과 같이 표현된다.

$$T(F_\varepsilon) = \mu(F_\varepsilon) + z_\alpha \sigma(F_\varepsilon) \tag{2.7}$$

식(2.7)에서와 같이 섭동된 범함수 $T(F_\varepsilon)$는 $\mu(F)$와 $\sigma(F)$의 섭동된 범함수 $\mu(F_\varepsilon)$과 $\sigma(F_\varepsilon)$의 선형결합으로 표현된다. 여기서 z_α는 하나의 x가 정규모집단에 추가되어도 거의 변하지 않는다. 즉, 섭동된 분포함수 F_ε에서도 α 분위수에는 큰 변화가 없다고 할 수 있다. 식(2.2)와 식(2.7)을 이용하여 α 분위수인 범함수 $T(F)$에 대한 영향함수를 유도하면 다음 식과 같다.

$$\begin{aligned} IF(x, \mu + z_\alpha \sigma) &= \lim_{\varepsilon \to 0} \frac{T(F_\varepsilon) - T(F)}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \frac{[\mu(F_\varepsilon) + z_\alpha \sigma(F_\varepsilon)] - [\mu(F) + z_\alpha \sigma(F)]}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \frac{[\mu(F_\varepsilon) - \mu(F)] + z_\alpha [\sigma(F_\varepsilon) - \sigma(F)]}{\varepsilon} \\ &= IF(x, \mu) + z_\alpha IF(x, \sigma) \\ &= x - \mu + z_\alpha \left\{ \frac{1}{2\sigma} [(x-\mu)^2 - \sigma^2] \right\} \end{aligned} \tag{2.8}$$

식(2.8)과 같이 $\mu + z_\alpha \sigma$의 영향함수는 μ와 σ의 각각의 영향함수 $IF(x, \mu)$와 $IF(x, \sigma)$로 표현된다. 즉, 범함수들의 선형결합에 의한 범함수의 영향함수도 각각의 범함수의 영향함수들의 선형결합으로 표현된다.

2.3 허용구간의 길이와 허용한계의 경험적 영향함수

가. 경험적 영향함수의 정의

범함수 $T(F)$는 모집단 분포함수 F에서 정의되는 모수 또는 모수의 함수로 정의된다. 이러한 모수 또는 모수의 함수를 표본으로부터 추정하는 추정량은 경험적 분포함수(empirical distribution function) F_n에 의해서 $T(F_n)$으로 정의된다. 따라서 추정량 $T(F_n)$ 역시 경험적 분포함수 F_n에서 정의되는 범함수라고 할 수 있다. 하나의 관찰치 x가 추정량에 대하여 갖는 영향을 측정하는 도구를 경험적 영향함수(empirical influence function)라고 한다(Cook & Weisberg, 1982). 추정량에 대한 경험적 영향함수는 범함수에 대한 영향함수식에서 모수 대신 모수의 추정량을 대치함으로써 쉽게 구할 수 있다. 또한 범함수에서 섭동인 ε은 추정량에서는 관찰치 한 개가 추가되는 것으로 간주되므로 $1/(n+1)$에 해당된다.

나. 허용구간의 길이와 허용한계에 대한 경험적 영향함수

관찰치 x가 양측 허용구간에 대하여 갖는 영향은 허용구간의 길이에 대하여 갖는 영향과 같다. 허용구간의 길이는 $2k_{\alpha/\gamma, n}s$이다. 각 관찰치가 허용구간의 길이에 미치는 영향은 표본표준편차 s에만 종속된다. 그러므로 s의 경험적 영향함수는 σ의 영향함수 식(2.6)을 이용하여 다음 식과 같이 표현된다.

$$EIF(x, s) = \frac{1}{2s}[(x-\bar{x})^2 - s^2] \qquad (2.9)$$

단측 허용구간은 어느 한쪽으로 양의 무한대 또는 음의 무한대이므로 관찰치의 영향을 구간의 길이로 표현할 수 없다. 따라서 이 경우에는 추정된 다른 한쪽의 한계에 대하여 관찰치가 갖는 영향을 측정한다. 상한의 허용한계 $\bar{x}+k_{\alpha/\gamma,n}s$ 에 대한 경험적 영향함수는 식(2.8)과 같은 원리로 평균 및 표준편차에 대한 경험적 영향함수를 이용하여 다음 식과 같이 나타낸다.

$$EIF(x, \bar{x}+k_{\alpha/\gamma,n}s) = (x-\bar{x}) + k_{\alpha/\gamma,n}\left\{\frac{1}{2s}[(x-\bar{x})^2 - s^2]\right\} \qquad (2.10)$$

3. 모의실험

3.1 모의실험 설정

앞에서 유도한 추정량 s와 허용한계 $\bar{x}+k_{\alpha/\gamma,n}s$ 에 대한 경험적 영향함수들의 수행능력을 평가하기 위해서 모의실험을 실시하였다. 정규분포 $N(10, 4^2)$에서 크기가 각각 10, 30, 50, 80, 100인 표본을 100번씩 반복하여 추출하였다. 추출한 표본을 대상으로 식(2.9)와 식(2.10)을 이용하여 표본의 각 관찰치에서의 영향함수 값을 계산하였다. 영향함수 식(2.9)와 식(2.10)에 각각 $1/(n+1)$을 곱하면 각 관찰치에 의한 표본표준편차 및 허용한계의 추정된 변화량이 된다. 이 추정된 변화량과 실제 변화량을 이용하여 Pearson의 표본상관계수(sample correlation coefficient)를 구하였다. 이때 실제 변화량은 전체 표본을 이용한 표본표준편차 및 허용한계와 각 관찰치를 추가했을 때의 표본표준편차 및 허용한계의 차이를 나타낸다. 또한 100번씩 반복 추출한 표본들로부터 구해진 상관계수들에 대한 평균 및 표준편차를 계산하여 유도한 영향함수의 타당성을 평가하였다.

3.2 모의실험 결과

표 3.1과 표 3.2는 크기가 30인 표본에 대한 모의실험 결과를 나타낸다. 표 3.1은 100번 반복 추출한 표본 중에서 임의의 한 경우에 해당되는 결과이다. 이 표본의 경우, $\bar{x}=9.7322$, $s=4.6030$ 이고 $\bar{x}+k_{0.95/0.95,30}s=19.9502$이다. 표 3.1에서는 표본평균, 표본표준편차 그리고 허용한계에 대한 각 관찰치의 영향함수값(IF), 영향함수로 추정한 통계치들의 변화량(EC), 그리고 각 관찰치를 추가했을 때의 실제 통계치들의 변화량(TCA)이 제시되었다. 또한 각 관찰치가 제거되었을 때의 통계치들의 변화량(TCD)도 추가로 제시하였다. 표본표준편차는 각 관찰치에서 표본표준편차의 추정된 변화량과 실제 변화량이 거의 차이가 없는 것으로 나타났고, 두 변화량에 대한 상관계수가 $r_1=0.9999364$ 로 나타나 두 변화량은 서로 선형관계를 보임을 알 수 있다. 허용한계에 대한 변화량의 차이는 표본평균의 영향함수값이 고려되었기 때문에 표본표준편차보다는 차이가 있지만 근소한 차이로 나타났고 상관계수는 $r_2=0.9992071$ 로 역시 선형관계를 보임을 알 수 있다. 이것은 유도한 영향함수가 바람직한 결과를 반영하고 있음을 의미한다.

표 3.1의 결과를 정밀히 검토하기 위하여 그림 3.1과 그림 3.2로 나타내었다. 그림 3.1은 각 관찰치들에 대한 세 통계량의 영향함수 값을 도시하고 있다. 표본평균에 대한 영향함수 값은 식(2.4)에

서 보듯이 관찰치에 대해서 직선경향을 보이고 평균과 거리가 먼 관찰치 일수록 영향값이 크게 나타나고 있다. 표본표준편차와 허용한계에 대한 영향함수는 동시에 관찰치에 대해서 이차식(quadratic)의 경향을 보이고 있다. 표본표준편차에 대한 영향함수는 약 10을 중심으로 대칭인 이차식을 보이고 있으며 특히 관찰치가 모집단의 표준편차를 고려한 6과 14 근처에서 영향함수의 값이 0으로 나타나고 있다. 또한 허용한계에 대한 영향함수도 약 7을 중심으로 3과 11 근처에서 영향함수의 값이 0으로 나타나고 있고, 표준편차에 대한 영향함수에 비해서 작은 값의 관찰치 보다는 큰 값의 관찰치에서 더 큰 영향함수의 값을 보이고 있다. 이것은 표본평균에 대한 영향함수와 $k_{.95/.95,30}$ 배한 표본표준편차의 영향함수가 더해진 효과 즉, 상한에 대한 영향함수이기 때문이다. 그림 3.2는 세 통계량에 대한 두 변화량을 도시화하였다. 전체적인 패턴이 그림 3.1에서의 패턴과 동일하게 나타났다. 평균 및 표준편차에서는 두 변화량이 거의 일치하였으나 허용한계에서는 두 변화량이 일정한 차이를 나타내고 있음을 볼 수 있다. 이것에 대해서는 심도 깊은 분석이 이루어져야 할 것이다.

표 3.2에서는 여러 크기의 표본을 100번씩 반복 추출하여 각 크기에서의 표본표준편차 및 허용한계에 대해서 영향함수로 추정한 변화량과 두 가지 방법으로 구한 실제 변화량간의 상관계수들의 평균과 표준편차를 나타낸다. 두 통계량의 경우 모두에서, 표본 크기가 커질수록 추정된 변화량과 추가방법으로 구한 실제변화량 사이의 평균상관계수는 1.0으로 근접하였고, 상관계수에 대한 표준편차도 표본크기가 커질수록 거의 0.0으로 근접하였다. 또한 제거방법으로 구한 실제변화량과의 결과에서도 거의 일치한 결과를 보이고 있다.

4. 결론

본 연구에서는 표준편차 σ의 영향함수를 이용하여 허용한계에 대한 영향함수를 유도하였다. 정규분포에서 추출한 확률표본들을 대상으로 각 관찰치가 표준편차와 허용한계의 추정량에 미치는 영향을 측정한 결과, 유도된 영향함수가 타당함을 확인하였다. 본 논문의 결과는 신뢰구간에 대해서도 동일한 방법으로 활용될 수 있을 것이다. 그리고 모의실험을 보완하여 영향함수 값에 대한 적당한 기준을 설정하면, 독립된 표본에 존재하는 이상치를 찾는 문제에도 활용할 수 있을 것이다.

표 3.1. 크기 30인 한 표본에 대한 모의실험 결과

id	x	Average				Standard Deviation				Tolerance Limit			
		IF	EC	TCA	TCD	IF	EC	TCA	TCD	IF	EC	TCA	TCD
1	17.31	7.58	0.24	0.24	0.26	3.94	0.13	0.12	0.15	16.32	0.53	0.46	0.54
2	7.44	-2.30	-0.07	-0.07	-0.08	-1.73	-0.06	-0.06	-0.06	-6.13	-0.20	-0.26	-0.27
3	12.91	3.17	0.10	0.10	0.11	-1.21	-0.04	-0.04	-0.04	0.49	0.02	-0.04	-0.04
4	10.70	0.97	0.03	0.03	0.03	-2.20	-0.07	-0.07	-0.08	-3.92	-0.13	-0.19	-0.20
5	14.06	4.33	0.14	0.14	0.15	-0.27	-0.01	-0.01	-0.01	3.73	0.12	0.06	0.08
6	3.19	-6.54	-0.21	-0.21	-0.23	2.35	0.08	0.07	0.09	-1.32	-0.04	-0.10	-0.08
7	18.93	9.19	0.30	0.30	0.32	6.88	0.22	0.21	0.26	24.47	0.79	0.72	0.85
8	11.03	1.30	0.04	0.04	0.04	-2.12	-0.07	-0.07	-0.07	-3.41	-0.11	-0.17	-0.18
9	15.04	5.31	0.17	0.17	0.18	0.76	0.02	0.02	0.03	7.01	0.23	0.17	0.19
10	8.79	-0.95	-0.03	-0.03	-0.03	-2.20	-0.07	-0.07	-0.08	-5.84	-0.19	-0.25	-0.26
11	1.42	-8.31	-0.27	-0.27	-0.29	5.20	0.17	0.16	0.20	3.24	0.10	0.04	0.10
12	10.88	1.15	0.04	0.04	0.04	-2.16	-0.07	-0.07	-0.08	-3.64	-0.12	-0.18	-0.19
13	9.89	0.15	0.00	0.00	0.01	-2.30	-0.07	-0.08	-0.08	-4.95	-0.16	-0.22	-0.23
14	6.17	-3.56	-0.11	-0.11	-0.12	-0.92	-0.03	-0.03	-0.03	-5.61	-0.18	-0.24	-0.25
15	6.82	-2.92	-0.09	-0.09	-0.10	-1.38	-0.04	-0.05	-0.05	-5.97	-0.19	-0.25	-0.27
16	6.59	-3.15	-0.10	-0.10	-0.11	-1.23	-0.04	-0.04	-0.04	-5.87	-0.19	-0.25	-0.26
17	4.62	-5.11	-0.16	-0.16	-0.18	0.53	0.02	0.01	0.02	-3.93	-0.13	-0.19	-0.18
18	8.93	-0.80	-0.03	-0.03	-0.03	-2.23	-0.07	-0.08	-0.08	-5.75	-0.19	-0.25	-0.26
19	8.42	-1.32	-0.04	-0.04	-0.05	-2.11	-0.07	-0.07	-0.07	-6.01	-0.19	-0.25	-0.27
20	5.72	-4.02	-0.13	-0.13	-0.14	-0.55	-0.02	-0.02	-0.02	-5.24	-0.17	-0.23	-0.24
21	10.66	0.93	0.03	0.03	0.03	-2.21	-0.07	-0.07	-0.08	-3.97	-0.13	-0.19	-0.20
22	15.10	5.37	0.17	0.17	0.19	0.83	0.03	0.02	0.03	7.20	0.23	0.17	0.20
23	7.78	-1.95	-0.06	-0.06	-0.07	-1.89	-0.06	-0.06	-0.07	-6.14	-0.20	-0.26	-0.27
24	15.09	5.36	0.17	0.17	0.18	0.81	0.03	0.02	0.03	7.16	0.23	0.17	0.20
25	6.99	-2.74	-0.09	-0.09	-0.09	-1.48	-0.05	-0.05	-0.05	-6.04	-0.19	-0.25	-0.27
26	3.26	-6.47	-0.21	-0.21	-0.22	2.24	0.07	0.07	0.09	-1.49	-0.05	-0.11	-0.09
27	12.21	2.48	0.08	0.08	0.09	-1.64	-0.05	-0.06	-0.06	-1.16	-0.04	-0.10	-0.10
28	14.57	4.84	0.16	0.16	0.17	0.24	0.01	0.01	0.01	5.37	0.17	0.11	0.14
29	2.62	-7.11	-0.23	-0.23	-0.25	3.19	0.10	0.10	0.12	-0.03	0.00	-0.06	-0.03
30	14.85	5.12	0.17	0.17	0.18	0.54	0.02	0.01	0.02	6.32	0.20	0.14	0.17

- IF : Influence Function, EC : Estimated Change using IF
- TCA : True Change when data is Added
- TCD : True Change when data is Deleted

표 3.2. 표본크기에 따른 영향함수의 평균상관계수 및 표준편차

Sample Size	Sigma				Tolerance Limit			
	Added		Deleted		Added		Deleted	
	AC	SD	AC	SD	AC	SD	AC	SD
10	0.9998	0.0001	0.9994	0.0004	0.9993	0.0008	0.9881	0.0099
20	0.9999	0.0001	0.9997	0.0002	0.9998	0.0003	0.9969	0.0024
30	0.9999	0.0001	0.9999	0.0001	0.9998	0.0002	0.9986	0.0010
50	1.0000	0.0000	0.9999	0.0001	0.9999	0.0001	0.9995	0.0003
80	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	0.9998	0.0001
100	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	0.9999	0.0001

- AC : Average of Correlation, SD : Standard Deviation

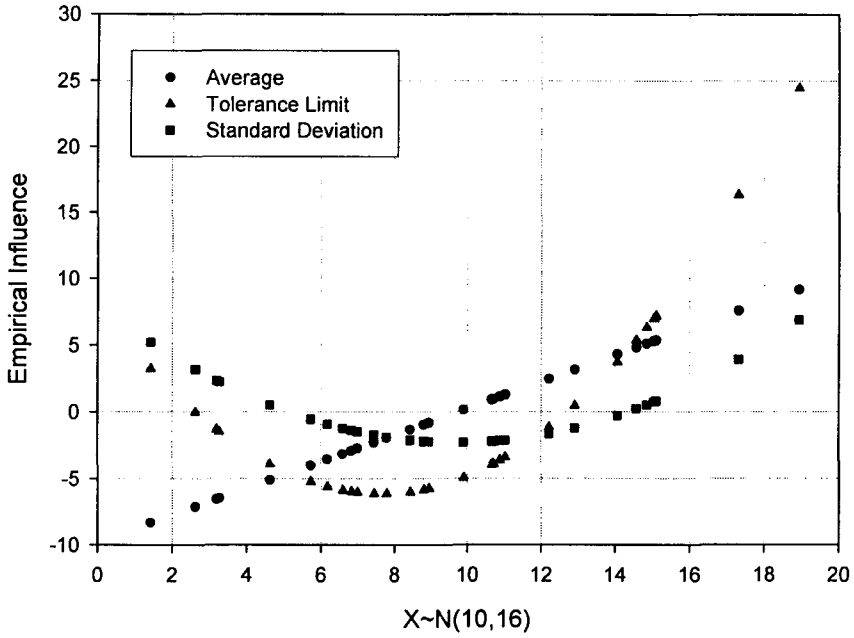

그림 3.1. 평균, 표준편차 및 허용한계의 경험적 영향함수.

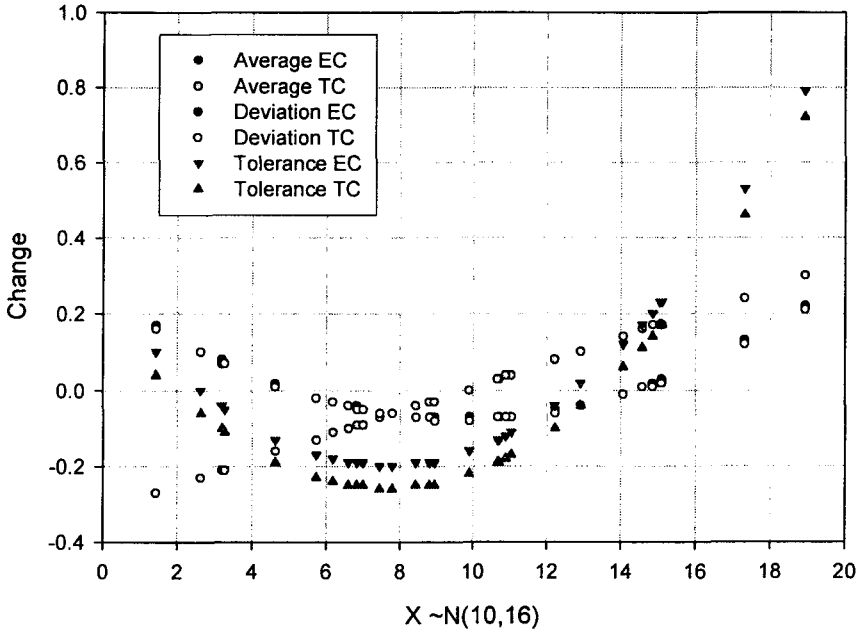

그림 3.2. 평균, 표준편차 및 허용한계에 대한 두 변화량

참고문헌

[1] Cook, R.D. and Weisberg, S.(1980). Characterization of an Empirical Influence Function for Detecting Influential Cases in Regression. Technometrics, Vol 22, No. 4, 495-508.

[2] Cook, R. D. and Weisberg, S.(1982). Residuals and Influence in Regression. Chapman and Hall Ltd.

[3] Dzieciolowski, K. and Ross, W. H.(1990). Assessing case influence on confidence intervals in nonlinear regression. The Canadian Journal of Statistics, Vol. 18, No. 2, 127-139.

[4] Guenther, W. C.(1972). Tolerance Intervals for Univariate Distribution. Naval Research Logistics Quarterly, Vol. 19, 309-333.

[5] Hampel, F. R.(1974). The Influence Curve and Its Role in Robust Estimation. Journal of the American Statistical Association, Vol. 69, No. 346, Theory and Methods Section, 383-393.

[6] Kim, H.(1992). Measures of Influence in Correspondence Analysis. The Journal of Statistical Computation and Simulation, Vol.40, 201-217

[7] Kim, H. and Lee, H.(1996). Influence Functions on χ^2 statistic in Contingency Tables, The Korean Communications in Statistics, Vol. 3, 69-76.

[8] Kim, H.(1998). A study on Cell Influences to Chi-square statistic in Contingency Tables, The Korean Communications in Statistics, Vol. 5, 35-42.

[9] Thomas, W.(1990). Influence on Confidence Regions for Regression Coefficients in Generalized Linear Models. Journal of the American Statistical Association, Vol. 85, No. 410, Theory and Methods, 393-397.

[10] Wald, A. and Wolfwiz, J.(1946). Tolerance Limits for a Normal Distribution. The Annals of Mathematical Statistics. Vol. 17, 208-215.

[2003년 3월 접수, 2003년 8월 채택]