■ 연구논문

망목특성에서의 자료분석을 통한 *SN*비의 선택[†]

임용빈
이화여대 통계학과

Selection of Signal-to-Noise Ratios through Simple Data Analysis[†]

Yong Bin Lim
Dept. of Statistics, Ewha Womans University

Abstract

For quality improvement, Taguchi emphasizes the reduction of variation of the quality characteristic. Taguchi has used the signal to noise ratios for achieving minimum dispersion of the quality characteristic with its location adjusted to some desired target value τ.

At each setting of design factors, the variance of the quality characteristic could be affected by the mean. In most cases, as the mean get larger, the variance tends to increase. The Taguchi's *SN* ratio corresponds to the case that the variance is proportional to the square of the mean. But the variance can increase faster or slower than the square of the mean. We propose to infer a linking relationship of the variance and mean through simple data analysis technique, and then use a reasonable *SN* ratio.

1. 서론

망목특성의 경우, 품질공학에서의 가장 중요한 목표는 품질특성치의 산포를 작게하면서, 평균을 목표치에 근접시키는 것이다. 품질특성치에 영향을 주는 요인들은 온도, 압력,

[†] 이 논문은 1992년 이화교수연구기금에 의해 지원을 받았음.

점토의 배합비 등과 같이 실험자가 그 값들을 제어할 수 있는 설계인자(design factors; x)와 각 가정의 전압의 불규칙성 등과 같은 제품의 사용환경조건, 원자재의 규격으로 부터의 약간의 차이, 기온, 습도 등과 같은 제조 또는 공정 주위 환경조건 등과 같이 실험자가 제어할 수 없는 인자인 잡음인자(noise factors; z)로 분류된다. 설계인자 중에서 제어에 비용이 많이 들어서 제어를 하지 않으면 잡음인자로 분류한다. 다구찌 박사의 가장 큰 업적은 잡음인자의 배치를 인위적으로 실험계획에 포함시킨 점이다. 다구찌 박사는 직교배열표를 이용하여 선택된 각각의 설계인자들의 실험조건에서 모든 잡음인자의 배치에 대한 실험을 실시하는, 설계인자의 배치와 잡음인자의 배치의 교적실험(product array)을 제안하였다. 제조공정 과정에서의 잡음인자(z)의 변화가 품질특성치의 변동을 초래한다. 평균을 목표치에 근접시키는 여러개의 설계인자들의 실험조건들이 존재하면, 그 중에서 잡음인자의 변화에 영향을 적게 받는, 즉 품질특성치의 산포를 작게하는 설계인자들의 조건을 찾는 것이 다구찌 방법의 기본 개념이다.

고전적인 실험계획법의 모형은 각각의 실험조건에서의 특성치의 산포는 동일하다고 가정하고, 특성치의 평균에 영향을 주는 인자를 찾는 문제에 관심을 가져왔다. 다구찌 박사는 설계인자와 잡음인자의 교호작용이 존재할 시에 실험조건마다 품질특성치의 산포가 같을 수가 없다는 사실에 근거하여, 품질향상을 위해서는 품질특성치의 산포에 영향을 주는 인자도 또한 알아낼 필요가 있다는 것을 잘 보여 주었다. 따라서 설계인자들 중에서 품질특성치의 산포와 평균에 영향을 주는 인자들을 알아내는 데에 관심이 있게 된다.

산포를 제어하는 인자를 찾기 위해서 다구찌(1987)는

$$SN = 10 \log \frac{\mu^2}{\sigma_y^2} \tag{1.1}$$

을 정의하고, 각각의 설계인자들의 실험조건에서의 SN비를

$$(SN)_i = 10 \log \frac{\bar{y}_i^2}{s_i^2} \tag{1.2}$$

으로 추정하였다. 여기서 $\bar{y}_i$와 s_i^2은 설계인자들의 i번째 실험조건의 n개의 잡음인자의 배치에서 얻어진 특성치 $y_{i1}, \cdots, y_{in}$의 자료평균과 자료분산이다. $(SN)_i$를 최대로 하는 실험조건이 산포를 가장 작게 하는 조건이고, 나머지 설계인자들 중에서 평균에 영향을 주는 인자들을 찾아내어 평균치를 목표치에 근접시키는 이단계 최적조건을 찾는 방법을 제시하였다.

SN의 통계적 성질들을 공부해 보도록 하자. 이차손실함수를 가정할 때에, 품질특성치 y의 설계인자들의 실험조건에서의 기대손실은

$$EL(y) = k \cdot E(y-\tau)^2 \quad (k\text{는 상수})$$

이 되어서, 기대손실을 최소로 하는 설계인자들의 조건을 찾는 문제는 품질특성치 y의 목표치 τ에 관한 평균제곱오차(MSE) 인

$$M = E(y-\tau)^2 = \sigma_y^2 + (\mu-\tau)^2$$

를 최소로 하는 설계인자들의 수준을 결정하는 문제와 동치이다.

특성치의 분산 σ_y^2이 평균 μ에 종속되어 $\sigma_y^2 \propto [f(\mu)]^2$로 주어지는 경우에는, 평균 μ를 τ에 일치시키는 조건과 분산 σ_y^2을 작게하는 조건을 분리해서 찾을 수가 없게 된다. 따라서 특성치의 분산 σ_y^2을 평균 μ에 종속된 부분인 $[f(\mu)]^2$와 평균에 독립된 부분에 해당되는 $\sigma_y^2/[f(\mu)]^2$의 곱인

$$\sigma_y^2 = \frac{\sigma_y^2}{[f(\mu)]^2} \cdot [f(\mu)]^2 \tag{1.3}$$

로 표현한다. 평균 μ가 목표치 τ에 근접한 실험조건에서는 식 (1.3)의 두번째 항인 $[f(\mu)]^2$는 $[f(\tau)]^2$로 근사된다. 따라서

$$\sigma_y^2 \simeq \frac{\sigma_y^2}{[f(\mu)]^2} \cdot [f(\tau)]^2 \tag{1.4}$$

로 근사되고, σ_y^2을 작게 하기 위해서는 식 (1.4)의 우변의 첫번째 항인 $\sigma_y^2/[f(\mu)]^2$를 최소로 하는 실험조건을 찾아야 한다. $\sigma_y^2/[f(\mu)]^2$는 평균에 종속되지 않는 성능측정량으로 Leon, Shoemaker & Kacker(1987)에서 정의된 PerMIA(Performance Measure Independent of Adjustment)와 동치이다. 예를 들면 분산 σ_y^2이 평균 μ의 제곱에 비례하는 경우에는, 평균이 목표치 τ의 근접한 실험조건에서는

$$\sigma_y^2 = \frac{\sigma_y^2}{\mu^2} \cdot \mu^2 \simeq \frac{\sigma_y^2}{\mu^2} \cdot \tau^2$$

이 되어 PerMIA는 σ_y^2/μ^2와 동치이고, PerMIA를 최소로 하는 실험조건은 다구찌가 제시한 $SN = 10\log(\mu^2/\sigma_y^2)$을 최대로 하는 실험조건이 된다.

실제문제에서 분산과 평균의 종속관계인 $\sigma_y^2/[f(\mu)]^2$는 일반적으로 알려져 있지 않다. 따라서 이 논문에서는 실험한 자료를 통해서 간단한 자료분석적 기법으로 종속관계를 추측하여 각각의 실험조건에서 $(SN)_i$를 선택하는 것을 제시하였다. 2절에서는 SN, PerMIA와 분산안정변환(variance stablizing transformation)이 간단히 요약되었고, 3절에서는 5개의 멱변환(power transformation) $Y = y^\lambda$, $\lambda = 1, .5, 0, -.5, -1$에 의해 변환된 자료들을 가지고, 각각의 설계인자들의 실험조건에서 계산된 자료평균과 자료표준편차의 산점도를 검토하여, 변환된 자료평균과 자료표준편차가 함수적으로 독립인 관계처럼

보이는 멱변환을 선택하여, 그에 대응되는 $(SN)_i$을 선택하는 것을 제시하였다. 또한 평균에 영향을 주는 인자를 찾는 성능특성치인 감도 $(S)_i$의 성질이 논의되었다. 4절에서는 3절에서 제시된 방법을 예제에 적용하여 분석하였다.

2. SN, PerMIA와 자료변환

다구찌는 통신공학에서 시스템의 신뢰성은 전달된 신호입력(signal input)의 힘과 잡음의 영향이 주는 힘의 비율(signal-to-noise ratio)로 표현된다는 기본개념을 이용하여 망목특성의 경우에 SN비를 식 (1.1)과 같이 정의하였다.

이절에서 소개하는 내용은 Leon, Shoemaker & Kacker (1987)와 Box (1988) 에서 연구된 결과를 정리하여 분산안정변환과 합리적인 SN비의 선택의 관점에서 고찰한 것이다.

평균 $\mu(x)$와 분산 σ_y^2의 관계가

$$P(d) = \frac{\sigma_y^2(x)}{[f(\mu(x))]^2} \tag{2.1}$$

로 주어진다고 하자. 여기서 $x=(d, a)$로 d는 설계인자들 중의 일부이고, 나머지 설계인자들에 해당되는 a로 평균의 조정이 가능하다고 가정하자. 설계인자 d의 수준의 값들이 주어졌을 때 평균 μ는 a만의 함수로 표현되고, $P(d)$는 a의 함수가 아니기 때문에 $P(d)$는 평균 μ와 독립이다. 따라서 $P(d)$는 산포특성치(a measure of dispersion)로 평균에 독립인 특성치가 되어서 PerMIA의 단조함수로 표현할 수 있어서 PerMIA와 동치이고, 평균제곱오차 MSE는

$$M(d, a) = P(d) \cdot [f(\mu(d, a))]^2 + (\mu(d, a)-\tau)^2 \tag{2.2}$$

이 된다. 평균 $\mu(d, a)$를 목표치 τ에 근접된 값으로 고정시켰을 때에 평균제곱오차 $M(d, a)$는 $P(d)$가 최소로 되는 d_0에서 최소로 된다. $d=d_0$에서 평균제곱오차 $M(d_0, a)$는 식 (2.2)에 의해 평균 $\mu(d_0, a)$의 함수이다. 따라서 $dM(d_0, a)/d\mu=0$으로 하는 μ_0에서 $M(d_0, a)$가 최소가 되고

$$\mu_0 = \tau - f(\mu_0)f'(\mu_0)P(d_0) \tag{2.3}$$

이다. $f(\mu)=\mu^\alpha$이 경우, $P(d)$는

$$P(d) = \frac{\sigma_y^2}{\mu^{2\alpha}}$$

이고, 식 (2.3)은

$$\mu_0 = \tau - \alpha \cdot \gamma_0^2 \cdot \mu_0$$

이 된다. 여기서 $\gamma_0 = \sigma_y / \mu_0$로 변동계수이다. 이 식을 정리하면

$$\mu_0 = \frac{\tau}{1 + \alpha \cdot \gamma_0^2}$$

가 된다. 따라서 평균 μ_0가 목표치 τ 보다 줄어 들은 값으로 결정되고, 평균조정인자 a를 $\mu(d_0, a_0) = \mu_0$가 되도록 a_0를 선택하면 (d_0, a_0)에서 평균제곱오차 $M(d, a)$가 최소로 된다. $\alpha = 1$ 경우, 즉 분산 $\sigma_y^2 \propto \mu^2$인 경우에는 $P(d) = \sigma_y^2(x)/\mu^2(x)$가 되고, 식 (1.1)의 SN과 비교할 때, $SN = -10 \log P(d)$가 된다. 따라서 $P(d)$의 최소화와 SN 의 최대화가 동치가 된다.

분산에서 평균에 종속되는 부분을 제거하는 또 다른 방법은 변환된 변수의 분산이 평균에 종속되지 않도록 변환을 결정하는 분산안정변환 (Variance Stabilizing Transformation)을 통해서 이다. 분산안정변환된 변수 $Y = h(y)$의 분산

$$\sigma_Y^2 \simeq [h'(\mu)]^2 \cdot \sigma_y^2$$

이 μ에 독립이라 가정하고, σ_Y^2이 설계변수 d에 의해서만 영향을 받는다고 하자. 그러면 평균제곱오차는

$$M(d, a) \simeq [h'(\mu(d, a))]^{-2} \cdot \sigma_Y^2(d) + (\mu(d, a) - \tau)^2 \qquad (2.4)$$

이 되어서 식 (2.2)와 비교할 때 평균에 독립인 성능특성치 $P(d)$의 최소화는 $\sigma_Y^2(d)$의 최소화와 동치가 되어서 산포에 영향을 주는 산포제어인자는 $\sigma_Y^2(d)$에 영향을 주는 인자들이다. 멱변환(Power Transformation)

$$Y = \begin{cases} y^\lambda & \text{if } \lambda \neq 0 \\ \log\ y & \text{if } \lambda = 0 \end{cases}$$

가 분산안정변환일 경우에는

$$\sigma_Y^2(d) \propto \mu^{2(\lambda-1)} \cdot \sigma_y^2$$

이 μ에 독립이고, $\sigma_y^2 \propto \mu^{2(1-\lambda)}$에 해당된다. 따라서 $P(d)$는

$$P(d) = \frac{\sigma_y^2}{\mu^{2(1-\lambda)}} \tag{2.5}$$

이 되고 합리적인 SN은

$$SN = 10 \log \frac{\mu^{2(1-\lambda)}}{\sigma_y^2} \tag{2.6}$$

이다.

3. SN 비의 선택과 감도의 성질

실제문제에서는 설계인자들의 실험조건에서의 분산과 평균과의 종속관계인 $[f(\mu)]^2$은 알려지지 않아서, 실험한 자료를 통해서 추측해야 한다. 설계인자 배치의 각각의 실험조건에서의 평균 μ_i와 분산 σ_i^2은 각각의 자료의 평균 $\bar{y}_i$와 자료의 분산 s_i^2으로 추정할 수 있다. Nair and Pregibon(1986)은 각각의 실험조건에서 ($\log \bar{y}_i$, $\log s_i^2$)의 산점도를 그려서 개략적인 기울기의 추정치가 2α인 경우 $\sigma_y^2 \propto \mu^{2\alpha}$에 해당되는 분산안정변환인 $Y = y^{1-\alpha}$을 적용한 후, 각각의 실험조건에서 변환된 변수의 표본분산 $s_{Y,i}^2$을 구해서 $\log s_{Y,i}^2$에 대한 분산분석에 의해서 산포제어인자를 찾는 방법을 제시하였다.

$Y = y^{\lambda^*}$가 분산안정변환이라면, 설계인자 배치의 각각의 실험조건에서의 변환된 자료의 평균 Y_i와 변환된 자료의 표준편차 $s_{Y,i}$의 산점도는 경향을 갖지 않는 랜덤한 구조를 가질 것이다. 이 경우의 합리적인 SN 비로는 식 (2.6)에 주어진 SN의 μ와 σ_y^2를 각각의 불편추정량인 원래 자료의 평균 $\bar{y}_i$, 자료의 분산 s_i^2로 대체한

$$(SN)_i = 10 \log \frac{\bar{y}_i^{2(1-\lambda^*)}}{s_i^2} \tag{3.1}$$

를 생각할 수 있다. 변수변환을 결정할 때에, 다구찌의 SN비에 익숙한 기술자들이 실제문제의 적용에 있어서 보다 합리적인 SN비를 선택하는 기법의 간편성이 중요하다. 따라서 우리는 λ^*를 결정하기 위해서 $[-1, 1]$을 등간격인 0.5로 나누어서, 5개의 멱변환(power transformation) $Y = y^{\lambda}$, $\lambda = 1, .5, 0, -.5, -1$ 된 자료들을 가지고, 각각의 설계인자들의 실험조건에서 계산된 자료평균과 자료표준편차의 산점도를 검토하여, 5개의 산점도 중에서 가장 랜덤하게 보이는 산점도를 선택하여서, 선택된 산점도에 해당되는 멱변환의 값을 λ^*의 값으로 선택하는 것을 제안한다. 산점도에 의한 구분이 명확하지 않을 경우에는 각각의 변환에 대한 자료평균과 자료표준편차의 표본상관계수를 계량적인 판단자료로 사용하여 표본상관계수가 0에 가까운 변환을 선택한다.

Box(1988)는 간단성(parsimony)과 분리성(seperation)의 원칙에 입각하여 자료의 평

균과 $\log s_{Y}^{2}$에 영향을 주는 인자들의 효과의 그래프인 lambda plot 를 동시에 비교하여서 적당한 분산안정변환을 선택하였다. 이 방법은 lambda plot의 작성과 그래프의 해석에 전문지식을 필요로 하기에 실용성이 문제가 된다. 본 논문의 방법은 lambda plot과 비교하여 별다른 전문지식이 없는 실무자들이 쉽게 적용할 수 있을 것이다.

평균조정인자를 찾기위한 성능특성치로 다구찌는 감도인

$$(S)_i = 10 \log (\bar{y}_i^2 - \frac{s_i^2}{n}) \tag{3.2}$$

를 때때로 사용하였다. 감도의 통계적인 성질들을 공부하도록 하자.

특성치 y의 분포가 정규분포라고 가정하자. Appendix의 Lemma 1에 의해 $\bar{y}_i^2 - s_i^2/n$은

$$E(\bar{y}_i^2 - \frac{s_i^2}{n}) = \mu_i^2$$

으로, i번째 설계인자들의 실험조건에서 모평균의 제곱의 불편추정치가 되고, 분산은

$$Var(\bar{y}_i^2 - \frac{s_i^2}{n}) = \frac{2}{n^2} (\frac{\sigma_i^4}{n-1} + \sigma_i^4 + 2n\mu_i^2\sigma_i^2)$$

으로, σ_i가 μ_i에 비례하는 경우에는

$$Var(\bar{y}_i^2 - \frac{s_i^2}{n}) \propto \mu_i^4 = [E(\bar{y}_i^2 - \frac{s_i^2}{n})]^2$$

이 되어서 분산이 기대값의 제곱에 비례하기 때문에 분산안정변환으로 log 변환을 한 후에 변환된 자료와 동치인 감도 $(S)_i$를 가지고 μ_i^2에 영향을 주는 평균조정인자를 찾는 방법이 합리적이다. 그런데 감도 $(S)_i$는 μ_i^2에 근거한 통계량이기 때문에 특성치가 양쪽 방향의 치우침의 정도와 같이 음의 값도 취할 수 있는 경우의 분석에는 적합하지가 않다.

감도 $(S)_i$는 각 실험조건에서의 분산을 고려하여 값이 결정된다. 즉, 똑 같은 자료의 평균인 $\bar{y}_i$의 값을 갖는 경우에도 그 실험조건에서의 자료의 분산의 값에 따라서 감도의 값이 조정되기 때문에, 단순한 자료평균 $\bar{y}_i$에 의한 분산분석 보다 직관적으로도 더 설득력이 있다.

4. 예

단순한 자료분석을 통한 SN 비의 적용예로서 박성현(1993)에서 소개된 자동차의 점화케이블의 코어 인장력을 규격 (40±15 파운드)에 맞도록 하는 제품을 개발하는 실험을 생

각하자. 자세한 실험자료는 박성현(1993)의 9.4절에 주어져 있으니 참고하시기를 바란다.

$\lambda=1$, 1/2, 0, $-1/2$, -1에서의 변환된 자료의 평균과 표준편차의 산점도를 비교해 보면 $\lambda=1$인 원래 자료의 평균과 표준편차의 산점도에서는 자료의 평균이 커짐에 따라서 표준편차가 증가하는 경향이 있고, $\lambda=1/2$과 $\lambda=0$의 산점도를 비교할 때 전자의 산점도가 후자의 경우 보다 상대적으로 더 랜덤한 구조를 보인다. $\lambda=-1/2$과 -1인 경우에는 평균이 커짐에 따라서 표준편차가 증가하는 경향을 보여 주기 때문에, $\sqrt{y}$ 변환이 분산안정변환임을 알 수가 있다. 산점도에 의한 비교와 더불어서 랜덤성의 계량적인 기준으로 각각의 변환된 자료의 평균과 표준편차의 표본상관계수를 구해보면, .54, $-.09$, $-.55$, .74, .85로 주어져서 $\lambda=1/2$에 해당되는 $\sqrt{y}$ 변환이 분산안정변환임을 뒷받침한다.

따라서 합리적인 SN비는 식 (3.1)의 $\lambda^*=1/2$인 경우인

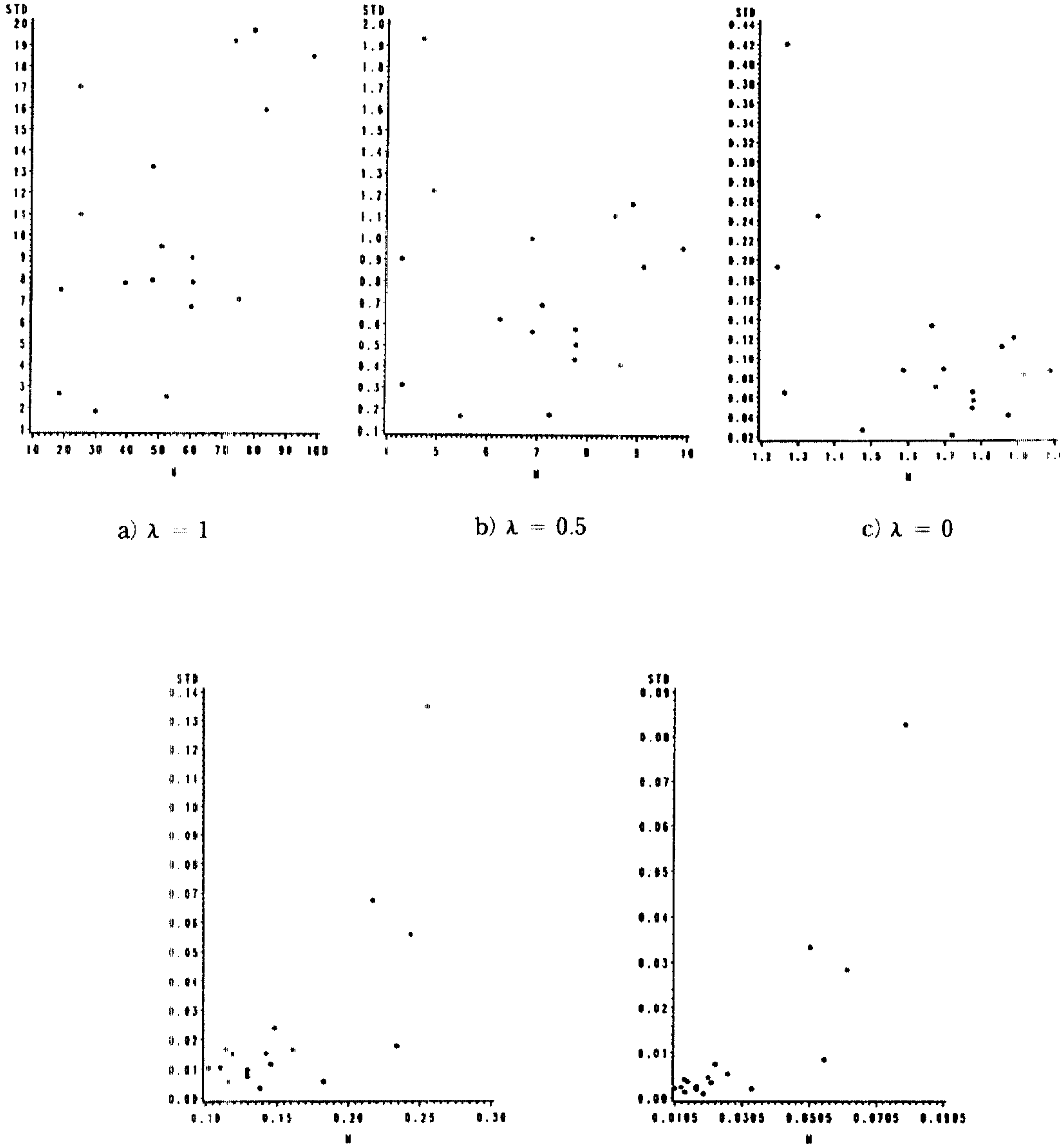
$$(SN)_i = 10 \log \frac{\bar{y}_i}{s_i^2} \tag{4.1}$$

이 된다. $(SN)_i$에 대한 분산분석표를 작성하면, C 인자에 대한 유의확률(p-value)이 0.9627로 매우 큰 값이 얻어져서 오차항으로 풀링할 수가 있고, 풀링한 후의 분산분석표가 표 1에 주어져 있다. 인자 B, H, D의 유의확률이 0.1 보다 작아서 산포제어인자로 분류할 수가 있고 인자 B의 유의확률이 0.0449로 가장 유의한 산포제어인자로 분류된다. $Y=\sqrt{y}$ 변환 후의 변환된 자료의 각각의 실험조건에서의 자료의 분산에 log를 취한 값인 log $s_{Y,i}^2$에 대한 분산분석표도 각 효과에 대한 유의확률이 식 (4.1)의 SN비에 의한 분산분석표인 표 1의 결과와 유사하여 거의 똑 같은 결과를 얻는다.

〈 표 1 〉 $(SN)_i = 10 \log (\bar{y}_i / s_i^2)$의 분산분석표

Source	DF	SS	MS	F value	Pr>F
A	1	25.61	25.61	3.01	0.1578
B	2	126.49	63.24	7.43	0.0449
D	2	85.45	42.72	5.02	0.0811
F	2	42.26	21.13	2.48	0.1990
G	2	48.38	24.19	2.84	0.1705
H	2	96.37	48.18	5.66	0.0681
I	2	56.61	28.30	3.33	0.1409
Error	4	34.03	8.51		
Corrected Total	17	515.18			

그런데 다구찌의 망목특성인 경우의 SN비에 의한 분산분석표에서는 I, B, D, H의 유의확률이 0.1 보다 작고, 인자 I의 유의확률이 0.0397로 가장 중요한 산포제어인자로 분류된다.



a) $\lambda = 1$ b) $\lambda = 0.5$ c) $\lambda = 0$

d) $\lambda = -0.5$ e) $\lambda = -1$

〈 그림 1 〉 5개의 역변환된 자료의 평균과 표준편차의 산점도

감도 $(S)_i$ 의 분산분석표는 박성현(1993)에서 분석한 민감도(sensitivity)

$$S_{n(i)} = 10\log(n\,\bar{y}_i^2) \tag{4.2}$$

의 분산분석표와 비슷한 결과를 얻는데, 유의확률이 0.8 보다 큰 인자 B와 H를 오차항에 풀링하면 인자 I에 대한 유의확률이 0.0339로 매우 유의하고, 인자 G가 유의확률이 0.1149로 평균에 영향을 줄 수도 있으리라 기대되기 때문에 인자 I, G를 평균조정인자로 분류할 수 있다.

식 (4.1)에 주어진 SN비와 감도 $(S)_i$ 에 관한 분산분석결과를 정리하면 다음의 결론을 내릴 수 있다.

산포제어인자 : B, H, D
평균조정인자 : I, G
기타제어인자 : A, C, F

공정의 최적조건을 찾는 의미에서 품질특성치에 약간의 영향을 줄 가능성이 있는 인자의 수준들을 실험자료에 근거하여 선택할 수도 있지만, 실제로 영향을 주는 인자들을 규명하는 것이 매우 중요하다. 그 이유는 통계적 분석에 의해서 알아낸 산포제어인자와 평균조정인자들이 왜 특성치의 산포와 평균에 영향을 주는 가를 공학의 전문지식을 이용하여 연구하면, 품질특성치를 향상시킬 수 있는 새로운 수준이나 또는 새로운 인자를 제시할 수도 있어서 현재의 최적조건 보다 더 나은 최적조건을 찾아줄 수 있는 이론적 근거를 제공하여 줄 수 있으므로, 통계적 분석과 공학적 이론의 Feedback 역할을 하여 더 좋은 품질의 제품을 설계할 수 있기 때문이다.

5. 요약

각각의 설계인자들의 실험조건에서 얻어지는 특성치들의 분산은 평균에 영향을 받는다. 많은 경우에 평균이 커짐에 따라서 분산이 커지는 경향이 있다. 다구찌가 산포제어인자를 찾기 위해서 제시한 SN비인 $(SN)_i = 10\log(\bar{y}_i^2/s_i^2)$은 분산이 평균의 제곱에 비례하여 커지는 경우이다. 그런데 분산이 평균의 제곱보다 더 느리게 또는 더 빠르게 커질 수도 있기 때문에 이 논문에서는 간단한 자료분석적 기법에 의해서 그 관계를 추측하여, 합당한 SN비를 사용할 것을 제시하였고, 평균조정인자를 찾기위한 통계량인 감도 $(S)_i$ 의 통계적 성질들을 논의하였다.

Appendix

Lemma 1. $y_1, \cdots, y_n$이 정규분포 $N(\mu, \sigma^2)$로 부터의 확률표본이고, $\bar{y}$와 s^2을 표본평균과 표본분산이라 하자. 통계량 $\bar{y}^2 - s^2/n$의 기대값과 분산은

$$E(\bar{y}^2 - \frac{s^2}{n}) = \mu^2,$$

$$Var(\bar{y}^2 - \frac{s^2}{n}) = \frac{2}{n^2}(\frac{\sigma^4}{n-1} + \sigma^4 + 2n\mu^2\sigma^2)$$

이다.

〈증명〉

$\bar{y}$의 분산은

$$Var(\bar{y}) = E(\bar{y}^2) - \mu^2$$

으로 표현할 수 있고, 따라서

$$\begin{aligned}\mu^2 &= E(\bar{y}^2) - Var(\bar{y}) \\ &= E(\bar{y}^2) - \frac{\sigma^2}{n} \\ &= E(\bar{y}^2 - \frac{s^2}{n})\end{aligned}$$

이다.

$\bar{y}$와 s^2은 서로 독립이기 때문에

$$Var(\bar{y}^2 - \frac{s^2}{n}) = Var(\bar{y}^2) + Var(\frac{s^2}{n}) \tag{A.1}$$

이다.

$\bar{y}$가 $N(\mu, \sigma^2/n)$을 따르기 때문에

$$E(\bar{y}^4) = \frac{3\sigma^4}{n^2} + 6\frac{\mu^2\sigma^2}{n} + \mu^4$$

이다. $\bar{y}^2$의 분산은

$$Var(\bar{y}^2) = E(\bar{y}^4) - (E(\bar{y}^2))^2$$
$$= 2\frac{\sigma^4}{n^2} + 4\frac{\mu^2\sigma^2}{n} \tag{A.2}$$

이고, $(n-1)s^2/\sigma^2$ 이 자유도 $(n-1)$ 인 카이제곱분포를 따르기 때문에

$$Var(s^2) = \frac{2}{n-1}\sigma^4 \tag{A.3}$$

이 된다. 식 (A.2)와 (A.3)을 (A.1)에 대입하면

$$Var(\bar{y}^2 - \frac{s^2}{n}) = \frac{2}{n^2}(\frac{\sigma^4}{n-1} + \sigma^4 + 2n\mu^2\sigma^2)$$

를 얻는다.

참고문헌

[1] 박성현(1993), 「품질공학」, 민영사.
[2] Box, G. E. P. (1988), "Signal-to-Noise Ratios, Performance Criteria and Transformations," *Technometrics*, Vol. 30, pp. 1-40.
[3] Leon, R. V., Shoemaker, A. C., and Kacker, R. N. (1987), "Performance Measure Independent of Adjustment: An Explaration and Extension of Taguchi's Signal to Noise Ratio," *Technometrics*, Vol. 29, pp. 253-285.
[4] Nair, V. N. (1992), "Taguchi's Parameter Design: A Panel Discussion," *Technometrics*, Vol. 34, pp. 127-161.
[5] Nair, V. N., and Pregibon, D. (1986), "A Data Analysis strategy for Quality Engineering Experiments," *AT & T Technical Journal*, pp. 73-84.
[6] Taguchi, G. (1987), *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost*, White Plains, NY: UNIPUB/Krau International.