

A Study on the Implementation of Experimental Design for the Analysis of Variability.

변동성분석을 위한 실험설계 모형과 실행절차에 대한 연구

A Study on the Modeling and Implementation of the Analysis of Variability

안해일*

*서경대학교 산업공학과 (hiahn@skuniv.ac.kr)

Abstract

Robust design of products or processes is a way to produce products or processes with less variability on condition that the product specification is met, so that the ratio of nonconforming products can be as minimized as possible. In such models as robust parameter design(RPD) originated by Taguchi and response surface methodology(RSM), the variances of noise variables are assumed to be measurable. However, there are circumstances in which the experimenters are unable to measure noise variables or perhaps are not aware of what the noise variables are. In this paper, it is demonstrated through an example that Taguchi's RPD approach can be conducted within the framework of the analysis of variability, with a view to a comparison of the two methods.

1. 서론

제품생산과정 또는 제조공정의 성능은 그 공정에서 생산된 부품 또는 제품의 특성치로 측정된 정확도(accuracy)와 정밀도(precision)로 표현되고 있다. 정확도와 정밀도는 위치-산포(location-dispersion)라고 하는 측도로 측정되며 통계학에서의 평균(mean)과 분산(variance)의 개념으로 설명될 수 있다고 할 수 있다.

최근에는 강건 파라미터 설계(robust parameter design)[10], 평균-분산 반응표면 모형(mean-variance modeling)[7], 변동성 분석 모형(analysis of variability model)[9], 일반화선형 모형(generalized linear model)[8, 11] 등 산포를 고려한 실험계획법 모형에 대한 연구가 진행되고 있는데 이는 제품 생산 공정의 산포가 부적합품의 발생과 직접적으로 연관되어 있다는 사실이 품질공학(quality engineering), 6시그마 설계(DFSS: design for six sigma) 등의 방법론과 함께 보급되면서 학계와 산업계 현장 실무자들이 새로이 관심을 갖게 되었기 때문이라고 생각된다.

제품 또는 공정설계에 있어 강건 파라미터 설계나 평균-분산 모형에서는 제품 생산 공정에 있어 제품의 품질 특성 치에 변동성(variability)을 유발하는 잡음인자(noise factor)와 이 잡음인자를 제어할 수 있는 제어인자(control factor)를 조사 분석하여 영향을 가장 적게 받도록 할 수 있는 제어인자의 수준을 찾아내고 평균이 특성치의 목표 값에 근접해야 한다는 제약 하에서 공정의 산포를 최소화 할 수 있는 이른바 최적조업조건을 알아내고자 하는

통계적인 실험 설계 및 분석 방법론이라고 할 수 있다.

여기서 강건(robust)이라는 단어는 잡음인자의 변화에 둔감(insensitive)하게 반응하여 잡음인자에 의하여 크게 좌우되지는 않는 다는 의미로 사용되고 있다. 또한 여기서 사용되고 있는 파라미터(parameter)의 의미도 수학에서 사용되는 매개변수(parameter)라든가 통계학에서 일반적으로 번역되는 모수(parameter)의 의미가 아니라 제어인자 또는 제어변수를 의미한다고 할 수 있다.

변동성분석(analysis of variability)에 대한 연구는 Bartlett and Kendall[1]에서와 같이 비교적 일찍이 시작된 것으로 알려지고 있으나 그동안 크게 주목을 받지 못하다가 80년대에 다구찌[10]의 강건 파라미터 설계 방법론이 알려지면서 또한 90년대에 DFSS(Design For Six Sigma)와 같은 관리기법이 알려지면서 다시 주목을 받기 시작한 것으로 생각된다.[2, 3, 11]

변동성 분석모형은 요인실험 설계나 다항 회귀 모형의 두 가지 형태를 모두 취할 수 있다. 변동성 분석모형이 강건 설계 모형과 다른 점은 모형에 잡음변수(인자)를 삽입하지 않고 있다는 점이다. 즉 강건 설계 모형에서는 외측배열(outer array)이 없이 반복실험(replication) 자료만을 사용한 경우라고 할 수 있으며 반응표면 계획법의 평균-분산(mean-variance) 모형에서는 잡음변수 항이 없는 다항회귀 모형에 해당한다고 할 수 있다.

본 고에서는 우선 다구찌의 품질공학에서의 강건 파라미터 설계 방법론을 살펴본 후 변동성 모형의 일반적인 모형화(modeling) 기법을 살펴보고 두 가지 방식을 비교하여 모형의 특징과 장단점에 대하여 논하고자한다. 또한 가중 최소자승법(weighted least squares) 또는 일반화 선형 모형(GLM: generalized linear model)의 접근방법[11]에 대하여서도 살펴봄으로써 변동성 분석 모형들 간의 유사점과 상이점 등에 대한 이해의 폭을 넓히는 데 목적이 있다.

2. 요인실험과 반응표면 모형

2.1 요인실험

요인실험분석은 흔히 2수준 계열과 3수준 계열로 나누어 생각하며 2수준 계열은 중심점이 추가되면 중심합성 계획법(CCD: central composite design)으로 발전되며 3수준계열은 박스-벤켄(BBD: Box-Behnken Design)이라는 명칭으로 발전되어 있는 것으로 보

여 진다.

요인실험 분석모형에 있어서는 분산분석에서의 자료구조 모형은 물론 반응표면 계획법에서의 다항회귀 모형도 사용될 수 있다. 외형적으로는 서로 다르게 보일 수 있으나 사실상 일대일의 변환이 가능하기 때문에 내용면에서는 동일한 분석기법이라고 할 수 있다.[7] 요인실험과 반응표면 모형의 동질성을 논하기 위해 다음과같은 예제를 우선 살펴 보자.

2.2 변동성분석 모형의 특징

다구찌의 강건 파라미터 설계(RPD) 적용사례 중 펌프실린더(pump cylinder) 내벽 마모실험(마모특성)에 대한 것이 문헌[10]에 존재한다. 내측배열을 2수준 $L_8(2^7)$ 으로 하였고 인자로서는 다음과 같은 것이 중요할 것으로 판단되었다.

- A : 주 재질(Material) - 경도
- B : 비중(Weight) - 무게/부피
- C : 표면 거칠기(Surface Roughness)
- D : 틈새(Clearance) - 거리
- E : 내벽 재질(Side Material) - 경도

구하고자 하는 요인효과로는 각 인자(A,B,C,D,E)의 주 효과와 $A \times B$, $A \times C$ 교호작용 효과이다. 마모자료는 실린더 내벽의 서로 다른 8곳(R_1, \dots, R_8)에서 측정된 자료이다.

품질 아카데미 연구소 논문집, 제7호, 2004년 4월, Seokyeong University
 Journal of Institute of Quality Academy, Vol.7, No.4, 2004, and
 and 같이 생각할 수 있다.

$$y = a_i + b_j + c_k + d_l + (ab)_{ij} + (ac)_{ik} + (abc)_{ijk} + \dots + \epsilon_{ijklm} \quad (1)$$

$i, j, k, l = -1, +1 \quad m = 1, 2, \dots, 8$
 $\epsilon_{ijklm} \sim N(0, \sigma^2)$

흔히 오차 항은 독립성과 등분산성을 가정한다. 주 효과(main effect)와 교호작용 효과(interaction effect)등 요인효과는 대비(contrast)로 계산되는데 각각 A, B, C, D, AB, AC 등으로 표기하기로 하면 정상적인 요인실험 분석순서에 따라 해당되는 변동도 계산할 수 있으며 분산분석표를 작성하여 유의성을 검정 할 수도 있다.

모든 인자(factor)가 연속형 변수로 생각할 수 있다면 이러한 계산결과를 이용하여 반응표면모형으로 변환할 수 있으며 다음과 같은 다항회귀 식으로 표현할 수 있다.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \dots + \epsilon \quad (2)$$

$x_1, x_2, x_3 \in [-1, +1]$

모수(계수)의 추정에 있어서는 내재하는 분포에 따라 즉 확률변수 이 어떠한 분포를 갖는가에 따라 다소 달라질 수 있지만 최소승법(least squares)과 최우 추정법(maximum likelihood estimation) 등 두 가지 추정방법이 사용될 수 있다.

수준수가 단 두 개인 경우에는 앞서 언급한 요

	A	B	A x B	C	A x C	D	E	마모자료(μm)								S/N	\bar{y}	s^2	s	ln(s)	
								R1	R2	R3	R4	R5	R6	R7	R8						
1	+1	+1	+1	+1	+1	+1	+1	12	12	10	13	3	3	16	20	-21.8717	11.125	34.411	5.8661	1.76918	
2	+1	+1	+1	-1	-1	-1	-1	6	10	3	5	3	4	20	18	-20.6023	8.625	46.268	6.8020	1.91722	
3	+1	-1	-1	+1	+1	-1	-1	9	10	5	4	2	1	3	2	-14.7712	4.500	11.143	3.3381	1.20540	
4	+1	-1	-1	-1	-1	+1	+1	8	8	5	4	3	4	9	9	-16.4836	6.250	6.214	2.4928	0.91343	
5	-1	+1	-1	+1	-1	+1	-1	16	14	8	8	3	2	20	33	-24.1539	13.000	104.286	10.2120	3.2357	
6	-1	+1	-1	-1	+1	-1	+1	18	26	4	2	3	3	7	10	-21.7136	9.125	74.411	8.6262	2.15480	
7	-1	-1	+1	+1	-1	-1	+1	14	22	7	5	3	4	19	21	-22.9584	11.875	64.696	8.0434	2.08485	
8	-1	-1	+1	-1	+1	+1	-1	16	13	5	4	11	4	14	30	-23.2710	12.125	74.696	8.6427	2.15672	
평균								12	14	6	6	4	3	14	18		9.5781	52.0156			
현재상태의 마모량								17	22	7	12	10	8	18	25		-24.1476	14.875	44.125		

표 2 펌프의 마모실험(정특성)

측정위치를 제어하기가 쉽고 또 그렇게 할 수 있다면 측정위치는 또 하나의 변량이 될 수 있다. 이러한 경우 측정위치를 블록(block) 또는 공변량(covariate)으로 간주하여 총 변동에서 공변량에 의한 변동을 제거한 후 요인효과를 비교분석하여야 한다. 이것은 난괴법(randomized block design)에서 흔히 볼 수 있는 분석방법이다.

실제로 2수준 계(2^m) 실험계획 모형을 적용시키면 적합도 결여가 유의한 것으로 나타나며 이상치로 보이는 측정치가 여러 곳에서(24, 40, 42, 64 번째 등)에서 발견된다. 결정계수 R²는 20%대에 머물고 있다. 하지만 블록인자를 포함시키면 결정계수 R²를 70%이상으로 높일 수 있었다.

한편 측정위치를 제어하기가 어렵고 측정위치가 매 측정 때마다 전혀 우연에 의해서 결정되는 상황이라면 가급적 산포 또는 변동성이 작은 조업수준을 알아 내보고자 하는 노력이 필요함은 당연하다 하겠다. 여기서 변동성 분석이 요구된다고 할 수 있다.

2.3 모형 간의 변환

인효과 대비를 이용하여 쉽게 추정할 수 있으며 추정된 결과는 A, B, C, D를 각인자의 주 효과 AB, AC를 두인자의 교호작용 효과라 할 때 다음과 같이 요약할 수 있다.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \hat{\beta}_{12} x_1 x_2 + \hat{\beta}_{13} x_1 x_3 + \dots + \epsilon \quad (3)$$

$x_1, x_2, x_3 \in [-1, +1]$

여기서 $\hat{\beta}_0 = \bar{y}$ 가 되는 이유는 반응표면 모형이 중심(centered)모형이기 때문이다. 다른 모든 계수 들을 계산할 때 요인효과를 2로 나누어 주는 이유는 요인효과는 높은 수준(+의 평균)에서 낮은 수준(-의 평균)을 제한 양이기 때문에 2단위 변화에 대한 차이인데 반해 다항식의 계수는 한 단위 변화에 대한 차이를 나타내기 때문이라고 할 수 있다.

그 외에도 변수의 제곱 항에 대한 곡률(curvature)변동 등이 포함될 수 있다. 자세한 내용은 Mont- gomery[5]와 Myers and Montgomery[6] 등에 설명되고 있다. 이와 같이 요인실험 모형과 반응표면 모형은 외형적 요소만 조금 다를 뿐 실제로는 동일한 내용이라고 할 수 있다.

A Study on the Implementation of Experimental Design for the Analysis of Variability.

3. 강건 설계 모형

3.1 SN비와 실험배치

손실함수의 가정으로부터 출발한다. 손실함수의 개념과 내용에 대한 설명은 많은 문헌[1-6]에서 설명하고 있다. 다구찌의 손실개념은 제품의 품질특성치가 공차역을 벗어나는 곳이 기점이 아니라 목표치로부터 벗어나는 기점부터 발생한다고 보는 특징이 있다. 따라서 목표치로부터의 편차의 제곱과 분산을 합한 통계량에 오메가(Ω) 변환이라고 하는 변환을 하여 SN비(signal-to-noise ratio)라는 측도를 창안하였다.

3.1.1 망목특성(nominal is best: NB)

실험 데이터로부터 SN비를 추정 할 때 다구찌는 잡음의 힘에 대한 신호의 힘으로 계산하였다. 구체적으로는 모평균(μ)의 제곱(μ^2)의 추정 값을 분산(σ^2)의 추정 값으로 나누어 구한다. 만약 n 개의 데이터 y_1, y_2, \dots, y_n 가 있다면 σ^2 의 불편추정치 값은

$$\hat{\sigma}^2 = V = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (4)$$

이 된다. $\hat{\mu}^2$ 은 $S_m = n\bar{y}^2 = (y_1 + y_2 + \dots + y_n)^2/n$ 인 경우에 $E(S_m) = \sigma^2 + n\mu^2$ 이므로 $S_m = \hat{\sigma}^2 + n\hat{\mu}^2$ 이며 $\hat{\mu}^2 = \frac{1}{n}(S_m - V)$ 이 된다. 따라서

$$\frac{\hat{\mu}^2}{\hat{\sigma}^2} = \frac{(S_m - V)}{nV} \quad (5)$$

를 얻는다. 실제로는 여기에 상용대수를 취하고 10을 곱하여 데시벨(decibel)단위로 나타낸 후 이 값을 최대로 하는 수준조합을 찾는 것이 실험설계 분석의 목적이 된다. 즉

$$SN_{NBZ} = 10 \log_{10} \frac{\hat{\mu}^2}{\hat{\sigma}^2} = 10 \log_{10} \left[\frac{(S_m - V)}{nV} \right] \\ = 10 \log_{10} \frac{(\bar{y})^2 - \frac{s^2}{n}}{s^2} \quad (6)$$

여기서 n 이 비교적 크다면 분자의 마지막 항은 무시되어 다음과 같이 될 수 있다.

$$SN_{NBZ} = 10 \log_{10} \frac{(\bar{y})^2}{s^2} \quad (7)$$

상용대수는 단조증가(monotonically increasing)함수라는 사실을 상기해야한다. 목표치(μ)가 정해져 있는 경우에는 고정된 값이므로 μ^2 을 추정할 필요가 없다. 따라서

$$SN_{NBZ} = 10 \log_{10} \frac{1}{s^2} = -10 \log_{10} s^2 \quad (8)$$

과 같이 생각하기도 한다.

3.1.2 망소특성(smaller is the better: SB)

목표치를 $\mu=0$ 으로 볼 수 있는 경우로서 $SN_{SB} = -10 \log_{10} \left[\frac{1}{n} \sum_{i=1}^n y_i^2 \right] = -10 \log_{10} \left[(\bar{y})^2 + \left(\frac{n-1}{n} \right) s^2 \right]$ (9)

을 최대화함으로써 목표치 최소화를 달성하는 것으로 볼 수 있다.

3.1.3 망대특성(larger is the better: LB)

목표치(μ)가 가급적 큰 $\mu=\infty$ 로 볼 수 있으며

$$SN_{LB} = -10 \log_{10} \left[\frac{1}{n} \sum_{i=1}^n \frac{1}{y_i^2} \right] = 10 \log_{10} \frac{n}{\sum_{i=1}^n 1/y_i^2} \quad (10)$$

$$\leq 10 \log_{10} \sum_{i=1}^n \frac{y_i^2}{n} = 10 \log_{10} \left[(\bar{y})^2 + \left(\frac{n-1}{n} \right) s^2 \right] \quad (11)$$

여기서 부등호가 성립하는 이유는 조화평균(harmonic mean)이 산술평균(arithmetic mean)보다 작거나 같다는 사실로부터 얻어낼 수 있다.

3.1.4 S/N비의 적용시 효과

이 세 가지 경우 모두 S/N비 계산식에 평균과 분산이 깊이 관련되어 있으며 이 S/N비의 크기를 최대화함으로써 산포를 최소화하게 된다고 할 수 있다. 망목, 망소, 망대 등 세 가지 경우 모두 S/N비를 최대화 하는 조업조건 즉 제어인자의 수준을 구하면 목표치가 달성된다는 제약 하에서 산포의 크기를 최소화 할 수 있다. 이와 같이 다구찌의 산포의 개념에는 평균과 분산이 모두 들어있다는 점이 특징이다.

3.1.5 S/N비 계산 예제

앞에서 주어진 계산 예제에 강건 파라미터 설계를 적용하기 위해서는 다음과 같이 배치하였다. 구하고자 하는 요인효과로는 각 인자(A~E)의 주 효과와 A×B, A×C 교호작용 효과이다. 분석은 다음과 같은 과정을 따른다.

1. S/N비 계산과 단일 품질 특성에 의한 분산 분석(ANOVA)

2. 최적 조건을 찾고 현재의 조건과 비교하여 소득(gain) 계산

3. 손실비용의 계산

계산 시 근거자료: 어떤 지점에서건 마모가 200 μm 을 초과하면 시장에서 문제점(trouble)이 발생하고 약 8만원의 손실이 발생된다. 설계수명이 20년 이고 현재 연간 마모량이 28 μm 이면 최적조건에서 생산되었을 때 얼마의 비용이 절약되는가? 단, 연간 생산량은 20만대이다.

이 문제의 해석하기 위해서는 Minitab Release 14.2를 사용하여 해결하여 본다면 다음과 같은 순서로 메뉴를 차례로 선택한다. 즉 “통계분석>실험 계획법> Taguchi 설계>Taguchi 설계분석”의 순으로 메뉴를 선택한 출력결과를 요약하면 다음과 같다.

Main Effects Plot for S/N Ratios Interaction Plot for S/N Ratios

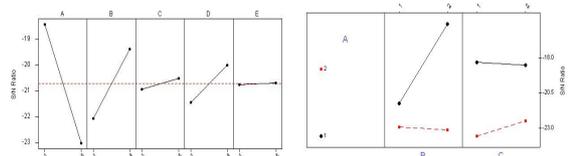


그림 1 주 효과와 교호작용 플롯(Minitab 결과)

Level	A	B	C	D	E
1	-18.4322	-22.0854	-20.9388	-21.4451	-20.7568
2	-23.0242	-19.3711	-20.5176	-20.0114	-20.6996
Delta	4.5921	2.7143	0.4212	1.4337	0.0572
Rank	1	2	4	3	5

표 3 반응 테이블(S/N비)

산포효과순위: $ABDCE$ 산포최소조건: $A_1B_2C_2D_2E_2$ 인자 A 와 B 사이 교호작용이 나타나고 있으나 최적 조건을 결정하는 데에는 변동이 없다. C 와 E 는 유의하지 않으므로 현재의 조건을 그대로 유지한다. 따라서 $A_1B_2D_2$ 를 최적 조건으로 선정하게 된다. 최소조건에서는 실험을 하지 않았으므로 확인실험을 필요로 한다. 실험결과 분석은 다음 과정을 포함한다.

1) 현재 조건의 S/N비

$$n = -10 \log_{10} \left[\frac{1}{n} \sum_{i=1}^n Y_i^2 \right]$$

$$= -10 \log_{10} \left[\frac{1}{8} (17^2 + 22^2 + 7^2 + 12^2 + 8^2 + 18^2 + 25^2) \right]$$

$$= -10 \log_{10} (259.875) = -24.1476 \quad (12)$$

2) 최적 조건의 S/N비

Minitab의 결과 최적 조건에서의 S/N비와 평균 분산에 대한 결과는 다음과 같다.

S/N Ratio	Mean	StdDev	Log(StdDev)
-14.9106	4.32813	2.86498	1.08434

$$-14.9106 = -10 \log_{10} (4.32813^2 + \frac{7}{8} \times 2.86498^2) \quad (13)$$

$$Gain = n_{\text{최적}} - n_{\text{현재}}$$

$$= -10 \log_{10} (MSD_{\text{최적}}) - [-10 \log_{10} (MSD_{\text{현재}})]$$

$$= -10 \log_{10} \left(\frac{MSD_{\text{최적}}}{MSD_{\text{현재}}} \right) = -14.91 - (-24.14) = 9.19 \quad (14)$$

$$\frac{(MSD_{\text{최적}})}{(MSD_{\text{현재}})} = 10^{\frac{-9.19}{-10}} = \frac{1}{10^{0.919}} = \frac{1}{8.28} \quad (15)$$

산포(MSD)는 8.28분의 1 정도로 줄어든 것으로 나타나고 있다. 이것을 기초로 손실비용과 공차설계를 하게 된다.

3.2 강건 설계 분석의 특징

다구찌 방식에서는 제어인자의 수준 조합에서의 분산이 서로 다를 수 있다는 가정 하에서 출발하기 때문에 일종의 변동성 분석모형으로 볼 수 있다. 교차배열을 사용하는 경우도 크게 다르지 않다. 분산 또는 평균을 추정하는 데에 있어서는 분산분석 기법을 그대로 적용한다. 하지만 분산분석을 적용하기 위해서는 관측치들의 등분산성이 보장되어야 한다. 다구찌의 방법에서는 자료의 구조 모형이 제시되지 않아 어떠한 가정 하에서 출발하고 있는지 알기 어렵다. 분산이 서로 다른 통계량에 상용대수를 취하고 S/N비를 구하면 근사적으로라도 등분산성이 달성된다는 사실이 통계적으로 입증된 바 없다.

또한 다구찌의 분석방법론에서는 목표치로부터 벗어난 평균과 분산이 S/N비라고 하는 하나의 측도로 통합되어 있어 반응치를 추정 후에 다시 역으로 평균반응과 분산반응을 분리할 수 없다. 평균반응은 불변인 것으로 가정하고 분산을 추정하기도 하지만 그렇다면 처음부터 평균과 분산을 분리해서 분석해도 된다고 해석할 수 있다고 생각된다.

한편 문헌에는 변동성분석 모형이라는 방법론이 있으며 평균(위치)과 분산(산포)을 분리하여 모형화하는 방법이 있다. 경우에 따라서는 직접 프로그램을 작성하여 실행해 볼 수도 있을 것이다. 비교적 최근에 SAS나 Minitab과 같은 통계패키지에서 변동성 분석 도구를 제시하고 있다. 이를 활용하여 변동성을 분석할 수 있는 방법도 생각해 볼 수 있다.

한편 문헌에는 변동성분석 모형이라는 방법론이 있으며 평균(위치)과 분산(산포)을 분리하여 모형화하는 방법이 있다. 경우에 따라서는 직접 프로그램을 작성하여 실행해 볼 수도 있을 것이다. 비교적 최근에 SAS나 Minitab과 같은 통계패키지에서 변동성 분석 도구를 제시하고 있다. 이를 활용하여 변동성을 분석할 수 있는 방법도 생각해 볼 수 있다.

4. 변동성 분석 모형

변동성 분석 모형은 평균뿐만 아니라 분산도 실험의 반응으로 생각한다는 점에 그 특징이 있다. 분산 모형화에 대한 연구는 Engle[2], Grego[3], Vining and Bohn[11]등에 의해서 수행되었는데 이는 시기적으로 보아 Ross[10]에서 볼 수 있듯이 다구찌의 연구결과가 이들에게 흥미를 유발 시킨 것이 아닌가 생각된다.

4.1 변동성 분석 모형의 공통적 특징

평균-분산모형에서의 분산은 제어가 사실상 불가능하다고 할 수 있는 잡음인자(noise variable)들의 산포로부터 전가된 때문 이기는 하지만 산포에 대한 분포에 대해서는 사전지식이 있어야만 한다는 것을 전제로 한다. 그러나 실제로 산업현장에서는 잡음인자를 측정하기 어려운 경우도 있을 수 있으며 경우에 따라서는 아직 잡음인자가 무엇인지조차 파악하지 못하고 있을 수도 있다.

잡음인자가 무엇인지 모르는 경우라도 즉 다시 말하면 제조공정 또는 생산된 제품의 산포의 원인이 무엇인지는 잘 모르지만 작업조건에 따라 균일하지 않은 위치효과(주효과) 또는 산포효과가 발생되고 있다는 사실을 인지하고 있는 경우에 적용할 수 있는 모형이 변동성 분석(analysis of variability) 모형이라고 할 수 있다.

변동성분석 모형의 기본 틀은 반응표면 계획법에서와 같이 다항회귀 모양을 갖고 있지만 일반적인 실험계획법 모형 특히 2수준 계열의 요인실험 계획법 모형과 다를 것이 없다. 다만 각 작업조건에서의 평균뿐만 아니라 분산(또는 표준편차)도 반응변수로 간주한다는 점이 다르다고 할 수 있다.

내용 면에서 변동성 분석모형이 강건 설계 모형과 다른 점은 모형에 잡음변수를 삽입하지 않고 있다는 점이다. 즉 다구찌 모형에서는 외측배열이 없이 반복실험(replication) 또는 반복측정(repetition) 자료만을 사용한 경우라고 할 수 있으며 반응표면 계획법의 평균-분산(mean-variance) 모형에서는 잡음변수 항이 없는 다항회귀모형에 해당한다고 할 수 있다.

외형 면에서 강건 설계모형과 다른 점은 다구찌 모형에서는 평균과 분산을 SN비라고 하는 단일 측도로 나타내고 있는 반면 변동성분석 모형에서는 반응이 평균(위치)과 분산(산포)으로 이원화(dual)되어 있어 모형 역시 이원화되어 있다고 하는 점이라고 할 수 있다.

4.2 분석 과정

반복실험 A Study of the Implications of the Experimental Design for the Analysis of Variance 라고 있다. 설계지점 \mathbf{x}_i 에서의 분산(s_i^2)에 관한 정보를 얻을 수 있는 경우에는 분산모형에서의 반응치 $\ln(s_i^2)$ 를 사용하여 분산의 추정모형을 만들 수 있다.

추정량 $s_1^2, s_2^2, \dots, s_d^2$ 들은 변량(확률변수)으로서 각 실험설계지점에서의 분산에 대한 불편 추정량이다. j 번째 설계지점에서의 반응값 $y_{1j}, y_{2j}, \dots, y_{nj}$ 이 정규분포를 따른다면

$$\frac{s_i^2(n-1)}{\sigma_i^2} \sim \chi_{n-1}^2 \quad (16)$$

이다. χ^2 분포는 감마분포의 특별한 경우로서 밀도함수에 지수함수가 들어 있는 지수함수 계급(exponential class)에 속한다.

일반적으로 지수함수 계급에 속하는 확률함수에는 베르누이(Bernoulli)분포, 이항(binomial)분포, 정규(normal)분포, 지수(exponential)분포, 감마(gamma)분포 등이 있다. χ^2 분포는 감마분포의 특별한 경우이므로 역시 지수함수 계급에 속한다. 지수함수 계급에 속하는 분포함수의 특징은 밀도함수로부터 위치(location)모수와 산포(dispersion)모수를 따로 분리해낼 수 있으며 분산은 평균의 함수형태로 주어진다. 이는 점이라고 할 수 있다. 자연대수를 취한 분산이 반응인 경우에는 내재하는 분포가 χ^2 분포로서 산포모수(dispersion parameter)가 $2/\nu$ 로 알려져 있는 감마분포이므로 엄밀히 말하면 감마분포 GLM(gamma GLM)에 해당한다고 할 수 있다[6]. 감마분포가 확실한 경우에는 gamma GLM 모형이 가장 효율적이라는 연구결과도 있다. 하지만 감마분포를 따른다는 사실을 입증하는 것도 간단한 문제는 아닌 것 같다.

여기서는 초기 단계의 변동성 분석 모형을 다루고 있으며 이러한 변동성 모형은 Bartlett and Kendall[1]에 의해 수립된 모형에 기초를 두고 있다고 알려지고 있다. 지수함수 계급(exponential class)에 속하는 확률변수에 자연대수(ln)를 취하면 점근적으로 정규분포를 따르는 확률변수를 얻을 수 있다. 자연 log함수인 $\ln(\cdot)$ 을 사용하는 이유가 거기에 있다.

4.2.1 산포(분산)모형

분산(산포) 예측모형은 흔히 대수-선형(log-linear) 모형이라고 불리는 것으로 다음과 같이 설명 될 수 있다. 우선 분산(산포)반응에 대한 모형으로서

$$\ln s_i^2 = \mathbf{x}_i' \boldsymbol{\gamma} + \epsilon_i \quad i=1,2,\dots,d \quad (17)$$

와 같이 수립한다. 이와 같이 분산 또는 표준 편차를 변환시키는 함수를 연계함수(link function)라고 부르며 내재하는 분포가 정규분포라고 할 수 없는 경우에도 적어도 점근적으로는 정규분포에 근사하도록 변환시킬 수 있다는 장점이 있다.

$\mathbf{x}_i, \boldsymbol{\gamma}$ 는 다중 회귀모형이라는 사실을 나타내고 있다. 여기서 $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_k)$ 이고 \mathbf{x}_i 는 다변량 다항식으로 설계행렬 X 의 i 번째 행이다. 또한 s_i^2 는 반복수 n 개의 관측치 들로부터 계산한 불편 표본 분산이며 실험설계지점 또는 수준조합의 수는 d 개가 있다고 가정하고 있다.

이모형은 오차가 비록 근사적이기는 하지만 정규분포를 따르며 분산도 어느 정도 균일하다고 할 수 있으므로 회귀모형의 기본 가정 즉 오차의 정규

4.2.2 평균(위치) 모형

평균(위치)반응에 대한 모형으로서 만일 오차들이 평균을 중심으로 정규분포를 한다는 가정 하에 다음과 같이 모형을 수립한다.

$$y_{ij} = \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_{ij} \quad i=1,2,\dots,d \quad j=1,2,\dots,n \quad (18)$$

여기서 $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_j)$ 이다. 또한 $\epsilon_{ij} \sim M(0,1)$ 이고 $\sigma_i^2 = \exp(\mathbf{x}_i' \boldsymbol{\gamma})$ 이다. 즉 분산이 독립적이기는 하지만 설계지점 \mathbf{x}_i 에 따라 그 크기가 다를 수 있다는 것을 나타내기 위한 식이다.

이 평균(위치) 모형에는 각 실험 설계지점(\mathbf{x}_i)에서의 분산이 서로 다를 수 있다는 가정 하에서 수립된 것이기 때문에 표준형 최소자승법(standard least square)을 사용할 수 없다. 오차 항들 간의 독립성을 가정할 수 있다면 적어도 가중 최소제곱법(weighted least square)을 적용할 수 있으나 독립성을 확신할 수 없는 경우에는 일반화 선형모형(GLM: generalized linear model)의 범주에 속하는 것으로 생각해야 한다고 할 수 있다.

4.2.3 가중 최소제곱법

일반화 선형모형(GLM) 중에서 가장 간단한 형태의 하나는 분산을 가중치로 사용한 경우로서 가중최소제곱법(weighted least square)의 문제로 볼 수도 있다.

다음과 같은 모형이 있다고 가정하여보자.

$$y_i = \mu(\mathbf{x}_i'; \boldsymbol{\beta}) + \epsilon_i \quad (i=1,2,\dots,n) \quad (19)$$

여기서 $\mu(\mathbf{x}_i'; \boldsymbol{\beta})$ 는 선형일 수도 있고 비선형일 수도 있다. 분산이 일정하지 않다는 가정 하에서는

$$Var(y_i) = Var(\epsilon_i) = v_i \quad (i=1,2,\dots,n) \quad (20)$$

이고 $v_i (i=1,2,\dots,n)$ 들이 사실상 알려져 있다면 $\boldsymbol{\beta}$ 의 추정은 다음의 최적화 문제로 볼 수 있다.

$$\text{Min} \sum_{i=1}^n \left[\frac{y_i - \mu(\mathbf{x}_i'; \boldsymbol{\beta})}{\sqrt{v_i}} \right]^2 \quad (21)$$

일반적으로 오차의 정규성을 가정할 수 없는 경우에는 GLM에서의 분산은 평균의 함수형태로 나타나기 때문에 단번에 추정할 수 있는 닫힌 형태(closed form)의 추정 식은 존재하지 않고 반복적으로 가중치를 다시 부여하여 수립될 때까지 반복하는 방법(iterative reweighted least squares: IRWLS)이 사용되고 있다[8]. 하지만 오차가 정규성(normality)을 갖고 독립적이라고 가정할 수 있다면 계수 $\boldsymbol{\beta}$ 와 분산 v_i 들은 독립적으로 추정될 수 있다.

이와 같이 변동성 분석모형에서는 평균모형과 산포모형 두 가지의 모형을 동시에 고려하여 최적의 조합조건을 찾아내는 것이 목표이다. 또한 이와 같은 대수-선형(log-linear)모형이 적용될 수 있는 경우는 매우 많으며 반복실험(replication)의 경우뿐만 아니라 반복측정(repetition)자료의 경우에도 용이하게 사용될 수 있어 Guage R&R의 문제에도 사용될 수 있다.

4.2.4 변동성분석 계산과정

일반적인 변동성 분석 절차는 다음과 같다고 할 수

$R(E_1) = e^{\text{A. Study on the Implementation of Experimental Design for the Analysis of Variability.}}$
 식은 다음과 같이 요약할 수 있다.

$$\widehat{\ln\sigma} = \widehat{\omega}_0 + \widehat{\omega}_1 X_1 + \widehat{\omega}_2 X_2 + \widehat{\omega}_{12} X_1 X_2 = 1.8244 - 0.3598 X_1 + 0.2231 X_2 + 0.1610 X_1 X_2 \quad (33)$$

여기서 $(\widehat{\omega}_0, \widehat{\omega}_1, \widehat{\omega}_2, \widehat{\omega}_{12}) = (1.8224, -0.3598, 0.2231, 0.1610)$ 이며 표준편차와 분산은 각각 다음 식들에 의하여 계산될 수 있다.

$$\widehat{\sigma} = e^{1.8244 - 0.3598 X_1 + 0.2231 X_2 + 0.1610 X_1 X_2} \quad (34)$$

$$\widehat{\sigma}^2 = e^{2(1.8244 - 0.3598 X_1 + 0.2231 X_2 + 0.1610 X_1 X_2)} \quad (35)$$

여기서 최적 수준 조합은 변동성이 가장 작은 작업조건으로서 위의 추정식을 편미분 하여 구한 방정식으로부터 구할 수 있다.

$$\frac{\partial \widehat{\ln\sigma}}{\partial X_1} = -0.3598 + 0.1610 X_2 = 0$$

$$\frac{\partial \widehat{\ln\sigma}}{\partial X_2} = 0.2231 + 0.1610 X_1 = 0 \quad (36)$$

$$X_1^* = -\frac{0.2231}{0.1610} = -1.3857, \quad X_2^* = \frac{0.3598}{0.1610} = 2.2348 \quad (37)$$

와 같이 계산 될 수 있으나 X_1 과 X_2 의 범위는 각각 $[-1, +1]$ 로 제한되어 있기 때문에 극소점이 제한영역 밖에 존재한다. 하지만 X_1 의 계수가 음수(-), X_2 의 계수가 양수(+)이기 때문에 X_1 의 높은 수준, X_2 의 낮은 수준, 즉 $X_1=+1, X_2=-1$ 에서 변동성이 가장 작은 결과를 얻을 수 있다. 여기서 한 가지 더 알 수 있는 것은 다음에 실험을 고려할 때는 A인자의 한 단계 더 높은 수준 B인자의 한 단계 더 낮은 수준에서의 실험을 설계해 보아야 할 것이라는 사실이다.

4.3.2 평균(위치)모형의 추정

앞에서도 언급한대로 평균반응을 추정하는 경우에는 각 계획지점에서의 분산을 가중치로 부여한 가중최소사승모형 또는 일반화 선형모형의 추정방법이 사용된다. Minitab으로 실행하기 위해서는 앞에서 추정된 표준 편차를 제공하여 분산의 추정치를 계산하고 이 분산의 역수를 가중치로 부여하여 평균 모형을 적합시켰으며 결과는 다음과 같다.

항	효과	계수	SE계수	T	P
상수		9.578	0.9015	10.62	0.000
A	-3.906	-1.953	0.9015	-2.17	0.035
B	1.781	0.891	0.9015	0.99	0.327
C	1.094	0.547	0.9015	0.61	0.547
AB	2.719	1.359	0.9015	1.51	0.137
AC	-0.719	-0.359	0.9015	-0.40	0.692
BC	2.094	1.047	0.9015	1.16	0.250
ABC	0.031	0.016	0.9015	0.02	0.986

표 8 Wear에 대한 추정된 효과 및 계수(평균모형)

R-제곱 = 25.88% R-제곱(수정) = 16.61%

항	효과	계수	SE계수	T	P
상수		8.880	0.8736	10.17	0.000
A	-5.417	-2.709	0.8736	-3.10	0.003

표 9 Wear에 대한 추정된 효과 및 계수

S = 1.02710 R-제곱 = 13.43% R-제곱(수정) = 12.03%

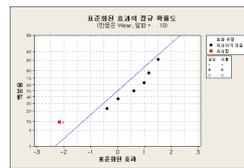


그림 4 정규확률도

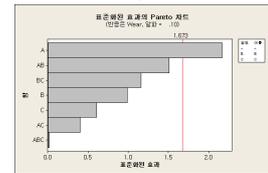


그림 5 파레토도

결과적으로 A인자를 제외하고는 유의적인 요인효과가 발견되지 않고 있다. 평균반응에 영향을 미치는 인자는 사실상 A인자가 유일하다고 할 수 있다. 따라서 추정된 평균반응 모형은

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 X_1 = \bar{y} + \widehat{\beta}_1 X_1 = 8.80 - 2.709 X_1 \quad (38)$$

와 같이 결론 내려질 수 있다. 산포와 평균으로 이원화된 모형은

$$\begin{cases} \widehat{\ln\sigma} = 1.8244 - 0.3598 X_1 + 0.2231 X_2 + 0.1610 X_1 X_2 & (\text{산포}) \\ \widehat{y} = 8.80 - 2.709 X_1 & (\text{평균}) \end{cases} \quad (39)$$

로 요약될 수 있다.

4.3.3 강건 설계와의 결과 비교

강건 설계에서의 최적조업조건은 $A_1 B_2 C_2 D_2 E_2$ 로 결정하였다. 여기서 $C_2 E_2$ 는 현재의 조업수준이다. D_2 는 변동성분석에서는 유의하지 않은 것으로 판단되었다. 한편 $A_1 B_2$ 는 변동성분석 모형에서는 $X_1=+1, X_2=-1$ 에 해당한다. (변환 시에 높은 수준과 낮은 수준이 바뀐 사실에 유념한다.) 현재의 조업조건 $A_2 B_1 C_2 D_2 E_2$ 은 $(X_1, X_2, X_3, X_4, X_5) = (-1, +1, -1, -1, -1)$ 에 해당한다.

강건 설계 분석 모형과 변동성분석 모형의 적용 결과는 매우 유사하며 크게 다르지 않은 것으로 보인다. 강건 설계분석에서는 D인자가 유의 하였지만 변동성분석 모형에서는 유의하지 않은 것으로 판단되었다는 점이 다소 다를 뿐이다.

하지만 강건 설계 모형에서는 평균과 분산의 효과가 분리되지 않아 모형으로부터 얻을 수 있는 정보가 제한적일 수밖에 없다.

예를 들자면 산포모형이나 평균모형을 적하시 결정계수는 약 20%정도에 지나지 않는다는 사실이 드러나고 있는 데 모형의 적합도의 면에서 모형은 매우 열등한 것으로 판단된다. 주된 이유는 모형실험 데이터에 6개 이상의 이상치(outlier)가 있는 것으로 추정되었는데 이러한 이상치들이 모형의 모수 추정치를 왜곡시키는 정도가 심하기 때문인 것으로 판단된다. 실제로 이러한 이상치가 발견될 때마다 제거하고 난괴법과 같은 블록인자를 도입하여 모형을 적합 시켜본 결과 70%이상의 결정계수를 얻을 수 있었다.

변동성 분석모형에서는 전통적인 분석방법론을 따르고 있고 분산과 평균을 분리하여 모형을 적용하고 분석하기 때문에 모형 적합과정에서의 모든 정보를 적나라하게 보여줄 수 있다. 한편 강건 설계분석에서는 이러한 이상치에 대한 탐지도 할 수 없을 뿐만 아니라 블록이나 공변량과 같은 외부 인자(회귀변수)를 도입하기도 어려우며 평균과 분산을 분리하지 않아 모형의 적합도 또는 적합도 결여 정도를 점검하기가 매우 어렵다.

변동성분석 모형의 또 다른 장점 중의 하나는 중심합성계획(central composite design: CCD)으로 확장하여 사용될 수 있다고 하는 점이다. 즉 평균 반

응 표면상의 극소점이나 극대점근처에 도달하면 품질 아카데믹 연구소 논문집 제 2권 2호 2004년 1월 15일
 대점이나 극소점의 위치를 정확히 알아내는 것이 Journal of Institute of Quality Academy, 인제대학교, 2004, 1월 15일
 중요하다 할 수 있다. 이것을 어느 정도 가능하게 하는 모형이 CCD이다. 산포모형이나 평균모형 모두 최적점에 거의 도달한 시기에는 CCD 설계를 해볼 수 있다. 제품 생산 공정에서 잡음변수가 확인되면 추가적으로 모형에 삽입하여 반응 표면의 강건 설계 평균-분산모형[7]으로의 확장도 가능하다

5. 결론 및 제언

본 연구에서는 최적 설계를 지향하는 실험계획법 모형들 중 변동성분석 모형에 관하여 모형의 특징과 실행절차 등에 대하여 살펴보았다. 변동성분석 모형은 잡음인자(변수)를 모형에서 포함시키지 않고 있다는 점에서 교차배열을 사용한 다구찌의 강건 파라미터 설계나 반응표면 계획법의 평균-분산 모형[7]과는 다르다고 할 수 있다. 평균-분산 모형에서 요구하는 자료는 잡음인자에 대한 종류, 분포 함수, 위치모수, 산포모수 등 상세한 정보를 요구한다. 따라서 제어가 어려운 인자를 제어한 상태에서의 실험이 요구되며 고가의 장비나 특수 설비가 필요하다고 할 수 있다.

현재의 생산설비나 적은 추가 비용으로 제조현장에서 바로 적용해 볼 수 있는 방법은 다구찌 강건 파라미터 설계 또는 변동성 분석이라고 할 수 있다. 하지만 다구찌의 방식은 통계적인 이론에 입각한 방법론이라고 보기 어렵다. 그래서인지 다구찌 기법에는 통계학적인 관점에서 미흡한 점이 많은 것으로 지적되곤 한다. 한편 변동성분석 모형은 통계학적인 관점에서도 매우 건전한 방법론이라고 생각된다. 특히 잡음인자(변수)의 존재 유무나 크기 등이 아직 파악되지 않은 상황에서는 다구찌의 강건 파라미터 설계와 동시에 변동성분석 모형을 적용해 볼 수 있다.

변동성 분석모형은 반응표면 계획법에서와 같이 모형이 다항회귀의 형태를 취하고 있다. 하지만 요인 실험계획 모형과 일대일 변환이 가능하기 때문에 사실상 동일한 분석모형으로 인식될 수 있다.

변동성분석 모형은 실행절차가 다구찌의 강건 파라미터 설계 모형에 비하여 다소 복잡한 것으로 보인다. 그 이유는 산포를 위한모형과 위치(평균)를 위한 모형으로 분리되어 이원화되어 있으며 또한 두 가지 모형 모두 일반화 선형모형(generalized linear model)의 범주에 속하기 때문에 일반적인 요인분석이나 회귀분석과는 모형에서의 기본 가정부터가 다르다고 할 수 있다. 예를 들어 산포모형에서는 내재된 분포 즉 잔차의 분포가 정규분포가 아닐 수 있다는 가정이 들어 있으며 평균 모형에서는 각 실험 계획지점에서의 산포가 독립적이긴하나 등분산성이 성립되지 않는다는 가정에서 출발하고 있다. 변동성분석 모형을 이해하고 산업현장에서 적용하기 위해서는 반응표면계획법(response surface method)내지는 일반화 선형 모형(generalized linear model)과 같은 분야의 전문지식을 필요로 한다.

어떠한 분석 모형을 산업현장에 적용하는 것이 좋은가 하는 문제는 어느 모형이 더 우수한가라는 문제를 생각하기 보다는 생산현장에서의 인력, 자료의 분석능력, 실험설비나 여건의 구비 정도에 따라 달라질 수 있는 것이 아닌가 생각된다.

이러한 관점에서 산업현장에서 분석모형의 도입을 몇 단계로 나누어 볼 수 있을 것이다. 첫째, 변

동성분석 모형의 중요성을 인식하고 산포의 제어를 위하여 관심을 갖아야 하며, 실험정신을 갖고 실험과 실험 계획의 중요성을 인식하는 단계가 필요하다고 생각된다. 둘째, 조업조건을 변경해가면서 실험하고 자료를 수집하여 강건 파라미터 설계와 같은 모형을 단순 모방해 보는 과정이 필요할 것으로 생각된다. 셋째, 다소 복잡한 통계기법이 요구되기는 하지만 변동성분석 모형과 같은 방법을 사용하여 평균과 산포를 분리 연구하는 단계도 필요할 것으로 생각된다. 넷째, 산포가 발생하는 원인을 찾아내고 잡음인자를 규명해내어 산포가 공정 또는 제품의 산포에 전가되는 메커니즘을 알아내는 단계로 접어 들 수 있을 것이다.

참고문헌

[1] Bartlett, M.S., and Kendall, D.G. (1946), "The Statistical Analysis of Variance Heterogeneity and the Logarithmic Transformation," Journal of the Royal Statistical Society, Series B, 8, 128-150.
 [2] Engle, J. (1992), "Modeling Variation in Industrial Experiments," Journal of the Royal Statistical Society Series C (Applied Statistics), Vol. 41, pp579-593.
 [3] Grego, J.M. (1993), "Generalized Linear Models and Process Variation," Journal of Quality Technology, Vol 25, pp 288-295.
 [4] Minitab (2004), Minitab Release 14 User's Manual.
 [5] Montgomery, D.C.(1991) Design and Analysis of Experiments, 3rd Edition, Wiley.
 [6] Myers, R.H., Montgomery, D.D., and Vining, G.G. (2002), General Linear Models with Applications in Engineering and the Sciences, John Wiley & Sons, New York.
 [7] Myers, R. H. and Montgomery, D. C.(2002), Response Surface Methodology, John Wiley & Sons Inc., 2nd Edition, New York.
 [8] Myers, R. H. and Montgomery, D. C. (1997), "A Tutorial on Generalized Linear Models", Journal of Quality Technology, Vol 29, No 3, pp274-291.
 [9] Nair, V.N. and Pregibon, D. (1988), "Analyzing Dispersion Effects From Replicated Factorial Experiments", Technometrics, Vol 30, No 3, pp 247-257.
 [10] Ross, P. J. (1989), Taguchi Techniques for Quality Engineering, McGraw-Hill.
 [11] Vining, G.G. and Bohn, L. (1998), "Response Surfaces for the Mean and the Process Variance Using a Non-parametric Approach," Journal of Quality Technology, Vol 30, pp 282-291.